Assessing Evaluative Measures for Internet-based Courseware

Kari J. E. Simpson, B. Ph. Ed

Department of Graduate and Undergraduate Studies in Education

Submitted in partial fulfillment of the requirements for the degree of Master of Education

Faculty of Education, Brock University

St. Catharines, Ontario

Abstract

Although there is a consensus in the literature on the many uses of the Internet in education, as well as the unique features of the Internet for presenting facts and information, there is no consensus on a standardized method for evaluating Internet-based courseware. Educators rarely have the opportunity to participate in the development of Internet-based courseware, yet they are encouraged to use the technology in their learning environments. This creates a need for summative evaluation methods for Internet-based health courseware.

The purpose of this study was to assess evaluative measures for Internet-based courseware. Specifically, two entities were evaluated within the study: a) the outcome of the Internet-based courseware, and b) the Internet-based courseware itself. To this end, the Web site www.bodymatters.com was evaluated using two different approaches by two different cohorts. The first approach was a performance appraisal by a group of end-users. A positive, statistically significant change in the students performance was observed due to the intervention of the Web site. The second approach was a product-oriented evaluation of the Web site with the use of a criterion-based checklist and an open-ended comments section.

The findings indicate that a summative, criterion-based evaluation is best completed by a multidisciplinary team. The findings also indicated that the two different cohorts reported different product-oriented appraisals of the Web site. The current research confirmed previous research that found that experts returning a poor evaluation of a Web site did not have a relationship to whether or not the end-users performance improved due to the intervention of the Web site.

Table of Contents

CHAPTER FIVE: SUMMARY, CONCLUSION, AND RECOMMENDATIONS

List of Tables

List of Figures

Glossary of Terms

End-user:

For the purpose of the current research, the term end-user will be used to describe the population for whom the Internet-based courseware was created; namely, grade seven students.

Expert:

For the purpose of the current research, the term expert will be used to describe the three adjudicators who responded to a questionnaire designed to elicit their formal education, informal education and employment experience where relevant to puberty education, pedagogy, or the Internet. These adjudicators were then given expert status based on the information gathered from the questionnaire.

Internet-based courseware:

This is a web site, or set of web sites designed to facilitate teaching and learning using the resources of the web.

CHAPTER ONE: THE PROBLEM

Introduction

There continues to be rapid and diverse growth of the Internet across a number of disciplines (Hoadley, 1996; Summers, 1996). In particular, the various levels of formal education have embraced the capabilities and the potential of the Internet for presenting facts and information in the pursuit of knowledge. The growing interest in the use of the Internet in education is evident in the proliferation of research journal articles. These articles describe various aspects of the use of the Internet in the learning environment (Flake, 1996; Hackbarth, 1997; Hannafin, Hall, Land, & Hill, 1994; Hoadley, 1996; Kearsly, Lynch, & Wizer, 1995; Maddux, 1994, Maddux & Johnson, 1997; Monahan & Dharm, 1995; Rankin, 1997; Shotsberger 1996; Staninger, 1994; Starr & Milheim, 1996). Current research describes the use of the Internet in the classroom, the unique features of the Internet, and the different methods of evaluating the applicability of the Internet (Kotecki & Siegel, 1998). Although there was consensus in the literature on the many uses of the Internet in education, as well as the unique features of the Internet for presenting facts and information, there was no consensus on a standardized method for evaluating Internet-based courseware (Reiser & Kegelmann, 1994).

Collis (1996) compared the introduction of the Internet in education to the earlier introduction of computers in education during the 1980s. The theme of Collis' comparison was that educators could gain knowledge from others' experience of introducing computers into education. One of the fundamental lessons was that the educator was the essential component of initiating the use of the Internet in education. The success of this initiative required educators to be partners in all stages of Internet curriculum development (Mauldin, 1996). Unfortunately, most educators did not participate in the developmental stages, but were required to use the technology to enhance specific aspects of their courses.

Therefore, this research project assessed different evaluative measures for

Internet-based courseware for the purpose of providing teachers with the knowledge required to assess Internet-based courseware.

The Problem Statement

The problem was twofold: a) The many uses of the Internet in education were well established in the related literature, yet there was no established standardized method for evaluating Internet-based courseware; and b) educators did not participate in the development of Internet-based courseware, yet they were encouraged to use it in their classrooms. These problems created the need for standardized, summative evaluation methods for Internet-based courseware that could be implemented by the educator.

The Purpose Statement

This project assessed evaluative measures for Internet-based courseware. Specifically, two entities were evaluated within this study: a) outcome of the product, and b) the product itself. The outcome referred to was the evaluation of learning performance by a target cohort. The product referred to was the Internet-based courseware that was evaluated by experts and end-users in independent sessions.

The Objectives of the Research

There was a need to develop guidelines to evaluate the end product of Internet-based courseware. To this end, the main objectives of this research were: a) to contribute to the knowledge which will be required to develop standardized Internet-based courseware evaluation guidelines, and b) to provide practical guidelines for educators to use to determine if they will incorporate Internet-based courseware into their learning environments.

Rationale and Importance of the Study

Although the volume of facts and information posted to the Internet continues to grow rapidly, no standardized evaluative process is used to filter this information. This is

a serious flaw in the application of the Internet as a tool in education. Without such a standardized review process, information which may be used in the learning environment lacks the due process necessary for a reliable, unbiased evaluation. A portion of this information is Internet-based courseware that is not subject to a rigorous review process. Educators rarely have the opportunity to participate in the developmental process of Internet-based courseware, and thus problems emerge for educators who are required to use the Internet in their learning environments.

The current research will be of use to practicing educators by providing a set of guidelines to determine if they will use a set of Internet-based courseware in their learning environments. The current research will also be of interest to researchers because it will contribute to the knowledge base that is required to develop sound assessment methods for summative evaluations of Internet-based courseware.

There were many ways to evaluate Internet-based courseware; however current related literature was inconclusive in establishing guidelines for its evaluation. To date, no guidelines for a formalized, standardized methodology exist to evaluate Internet-based courseware.

Scope and Limitations of the Study

The current research did not attempt to draw conclusions based on the broad assessments of the three evaluation methods conducted. Instead this work attempted to accomplish two tasks; a) to contribute to the knowledge required to develop standardized evaluation methods for Internet-based courseware, and b) to provide guidelines for educators to make practical use of the results of the study. The small populations used for the expert evaluation and the end-user evaluation limited the findings. The findings were also limited by the inherent problems of a criterion-based checklist, a quasi-experimental pretest and posttest design, and a content analysis of qualitative open-ended

comments. Therefore, the current research was a descriptive study of the results of different evaluative methods for the purpose of contributing to the knowledge required to develop standardized evaluation methods for Internet-based courseware.

Outline of Remainder of the Document

The remainder of the document includes a review of the related literature, a comprehensive outline of the pilot work that contributed to the development of the methodology, and the methodology itself. The document also describes the findings of the research and the interpretation of those findings. Finally, the conclusions, implications, and recommendations that were drawn from the research are provided.

The review of literature provides a discussion of the uses of the Internet in education, presents evaluation theory for Internet-based courseware, and provides the different evaluation methods for Internet-based courseware.

Chapter Three consists of a description of the research methodology and research design. The three pilot studies that were used to develop the research design are described and the limitations, sample, and data analysis are introduced.

Chapter Four introduces the findings of the research and provides an in-depth interpretation of the findings. The results of the expert, product-oriented evaluation and the end-user, product-oriented evaluation are provided and discussed, and a comparison between the findings of each are made. This protocol is repeated for the comments of the two cohorts. Finally, the end-user scores on the pretest and posttest performance appraisal are reported and discussed.

Chapter Five includes the conclusions that were drawn from the research and the implications for practice and theory, as well as recommendations for future research. The implications for practice include a list of guidelines to help educators evaluate Internet-based courseware, and the implications for theory include a description of how the

findings of this study can contribute to the knowledge required to create standardized evaluation methods for Internet-based courseware. The recommendations provide insight to guide future research.

CHAPTER TWO: REVIEW OF RELATED LITERATURE

Introduction

The current climate in education encourages educators to use information technology in their learning environments. In particular, the Internet is seen as a revolutionary way to provide learners with access to a broad spectrum of information, as well as to allow for communication capabilities that were not accessible even 5 years ago (Kearsley et al., 1995). The freedom to post information to the World Wide Web may be problematic for educators due to the unedited production of web sites (Symons, 1997). The majority of web sites were not reviewed or refereed by educational authorities (Wilkinson, Bennet, & Oliver, 1997). Wilkinson et al. (1997) explained that while educators were confident in the quality of material presented in textbooks because of the strict review processes that ensured quality, the educator does not have the same confidence with educational resources on the Internet.

Another inherent problem with Internet-based courseware was that the educator had no voice in the product and they were unable to change the material that was presented for their use. This problem suggested the need for educators to conduct summative evaluation of Internet-based courseware prior to introducing this medium into their classrooms.

Summary of Literature Reviewed

Uses of the Internet in Education

Educational web sites were designed to facilitate teaching and learning using the resources of the web (Ohl & Cates, 1997). Internet-based courseware was a set of learning material that was designed to further knowledge within a given domain.

Internet-based courseware enabled the learner to benefit from a number of educational opportunities. These opportunities were grouped into three categories: a) communication, b) information retrieval, and c) information sharing (Hackbarth, 1997).

Communication

The Internet supports a number of communication modalities that act as tools in the learning process. These tools include bulletin board systems (BBS), e-mail, conferencing capabilities, listservs, and newsgroups (Hackbarth, 1997; Lasarenko, 1997). The broad spectrum of communication tools available to students helps to dissolve the artificial wall between the classroom and the real world (Hackbarth, 1997; Hoadley, 1996).

A bulletin board system (BBS) is an electronic repository where messages could be posted to a virtual bulletin board and read at the convenience of the user (Holzschlag, 1997). Bulletin boards worked in asynchronous time and enabled the user to read a posted message, respond to it, or post their own message (Kearsley et al., 1995). Some bulletin board systems had editors who censored the messages that were posted in order to eliminate inappropriate messages.

E-mail is a method of communicating with individuals or groups in an asynchronous and time-efficient manner (Monahan & Dharm, 1995). E-mail could be used to link a class together for the purpose of using each other as support and it could be used to contact experts for class purposes (Hackbarth, 1997).

A chatroom was a cyberspace location where a synchronous, electronic discussion or conference could take place using chat software (Lasarenko, 1997).

Users who had Internet Relay Chat capabilities can join together to discuss issues in "real time". The primary benefit of synchronous communication was that the user did not have to wait for a reply, a benefit that facilitates the flow of the discussion. The primary drawback for synchronous communication, especially within the context of education, was that the educator could not control or censor the discussions to ensure that inappropriate discussions were avoided. This limitation could be avoided by meeting in a chatroom that was designated for a certain purpose and where all participants had a password to gain access.

A "listserv" was a program that managed e-mail discussion lists (Lasarenko, 1997). Listservs enabled students to subscribe to a daily mailing list to receive and transmit messages to members (Starr & Milheim, 1996). Each listserv disseminated information about a specific topic by transmitting messages about current issues, conferences, and concerns of the members. A listserv membership was an excellent method for a class to remain current in a specific area.

A newsgroup was a USENET discussion group. A USENET was an international collection of over 10,000 newsgroup discussion lists (Lasarenko, 1997).

Information retrieval

Information retrieval was a second category of educational opportunities that the Internet could offer students. The information that was available on the Internet was rich in both volume and variety (Monahan & Dharm, 1995). The Internet began as a repository for information by the American Department of Defense in the late 1960s (Starr & Milheim, 1996). It is now a network of databases that hold a bounty of facts and information. However, facts and information on their own do not

provide a rich educational experience. It was the pursuit of knowledge from using this wealth of information that would help the learner to develop their intellectual skills. Students were able to find their own solutions to tasks, to think critically about the information, to question the validity and relevance of web-based information, and to decide for themselves if the information they found answered their specific problems (Hannafin et al., 1994).

Another important aspect of the Internet was that it did not limit students to a time or a place. A student who had access to the Internet from their home or from their local library was no longer bound to the constraints of the school day, the classroom, or their schoolbooks. They could have access to this information at their convenience. The students could retrieve copies of class handouts that had been posted to the Internet, they could examine enrichment material at an address supplied by their instructor, they could look at answers to exercises, and use links to get to reference tools such as dictionaries or encyclopedias (Maddux & Johnson, 1997).

The information that was available to students could enable them to reach beyond the walls of the classroom by accessing the enormous collection of databases that the World Wide Web makes accessible (Hoadley, 1996; Starr & Milheim, 1996). Many programs had already been implemented in all levels of formal education to harness this vast collection of facts and information (Hackbarth, 1997; Kearsley et al., 1995; Starr & Milheim, 1996; Woolsey & Bellamy, 1997)

Information sharing

The final category of educational opportunities that the Internet could offer students was information sharing. Students were able to collaborate with others through posting their work, receiving feedback, collecting data, and working with students in other parts of the world (Hoadley, 1996; Kearsley et al., 1995; Rankin, 1997). The technology was changing rapidly to the point where asynchronous communication such as e-mail was no longer the only communication medium. There were concurrent communication measures in place through conferencing capabilities. With the ease of HTML and JavaScript, a student could post their ideas with graphics, film clips, and plans and could receive feedback from others who visit their web site (Hackbarth, 1997).

Hypertext and the Internet

> Constructivists enjoyed the possibility of the World Wide Web because it provided rich contexts, authentic tasks, collaboration for the development and evaluation of multiple perspectives, an abundance of tools to enhance communication and access to real-world examples and problems, reflective thinking, modeling of problem solving by experts in the content domain and apprenticeship mentoring relationships to guide learning. (Duffy & Bednar, 1992, p. 132).

Hypertext facilitated the rich contexts, the reflective thinking, and the problem solving challenges of Internet-based courseware. Hypertext was the language of the Internet. Hypertext was organized on the premise that all information had inherent interconnectivity and that the learner had the opportunity to follow their own path to

find their own answers (Staninger, 1994). Landow (1992) described hypertext as chunks of words or pictures that were connected electronically by numerous paths in an open-ended and continually incomplete environment. Landow (1992) indicated that the terms link, node, network, web, and path could best describe this open-ended and continually incomplete environment. The learning opportunities of hypertext were numerous. Hypertext allowed the students to create relationships between groups of facts (Hannafin et al., 1994). It helped students to develop associations between groups of facts which they can then assemble into knowledge (Staninger, 1994). Students were able to gain serendipitous information by navigating their own path to the information that they seek. It was the unique feature of the Internet that books and lectures could not provide, because hypertext had no functional beginning or end (Staninger, 1994). Learners needed to develop problem-solving skills and critical thinking skills to discern important information from irrelevant information (Hannafin et al., 1994). They needed to decide when they had acquired enough information and they learned new modes of information acquisition. The hypertext environment encouraged this type of learning.

Evaluation Theory for Internet-based Courseware

The existing literature on evaluating Internet-based curriculum was inconclusive. There were a number of authors that called for the implementation of formative evaluation strategies to be used to ensure quality; however, the typical educator did not have the opportunity to participate in the process of evaluation. Therefore, the need for sound summative evaluation methodology for educators was necessary. Through the summative evaluation process educators may decide to implement certain Internet-based curriculum into their classrooms.

Scriven (1967) coined the term formative evaluation as a way to describe the evaluation of educational programs during their development phase for the purpose of improvement. This concept transferred to the instructional design process, where authors were concerned with creating instruction to bring about changes in knowledge, skills, or attitude (Weston, McAlpine & Bordonaro, 1995). This type of evaluation was particularly important in the creation of Internet-based courseware. Formative evaluation was a dynamic process based on the systematic identification of areas needing modification throughout the design phase. It was based on continuous modifications for the purpose of guiding design and improving the final product. (Mauldin, 1996; Northrup, 1995; Weston et al., 1995). Formative evaluation was the optimal choice for Internet-based courseware development because it provided the opportunity for teachers and students to participate in concurrent evaluation of the Internet-based courseware (Mauldin, 1996).

Educators were typically expected to evaluate courseware as a final product. The expectation was that the courseware was either accepted or rejected for use. The term summative evaluation was central to the present study because educators did not always have the opportunity to participate in the formative evaluation of the courseware that they used in their classrooms. Summative evaluation was an evaluation method that provided comprehensive assessment of the final product (Persico, 1997). This type of evaluation was goal oriented and was used to ascertain whether or not the courseware had achieved the desired outcome (Flagg, 1994; Persico, 1997).

Harper (1988) introduced a fundamental principle, which underlies both

formative and summative evaluation, which could be applied to the review of Internet-based courseware. According to Harper, a multidimensional assessment that measures with more than one instrument should be used. The present study recognized the importance of a multidimensional model for summative evaluation and intended to assess three different methods of evaluation by two different cohorts.

Evaluation Methods for Internet-based Courseware

Education-technology literature presented many methods for evaluating computer-based courseware. These evaluation methods were grouped into two types of evaluation methods: product-oriented evaluation methods, which included observational evaluation, survey evaluation, and criterion-based checklist evaluations, and outcome-oriented evaluations such as performance appraisals. (Aedo, Catenazzi, & Diaz, 1996; Flagg, 1994; Lohr, Ross, & Morrisson, 1995; Neuman, & Marchionini, 1995; Reiser & Kegelmann, 1994). Other authors spoke to methods such as product-oriented evaluation, end-user evaluation, and criterion-based evaluation (Gros & Spector, 1994; Northrup, 1995; Sorge, Campbell, & Russell, 1993; Symons, 1997; Wilkinson et al., 1997). Each method was not mutually exclusive, and it was common to have different methods of evaluation sharing common features. Each method is discussed in the following paragraphs.

Performance evaluation

Reiser and Kegelmann (1994) suggested that evaluators measure what the students learned through gathering student performance data. The performance data eliminated the subjective bias of the expert and end-user evaluations. The objective of a performance evaluation was to compile empirical evidence of student

performance for the purpose of determining the instructional effectiveness of the courseware (Muller, 1985). Jolicoeur and Berger (1988) suggested that courseware evaluation should include a pretest and posttest of the student's performance to obtain the empirical data.

Northrup (1995) suggested selecting users of various genders, cultures, and races from the target audience to answer a prototype assessment tool. Northrup indicated that the assessment tool should be based on the objectives selected during the development of the instructional product. The assessment tool could be on line or paper and pencil and the responses of the performance data could guide revisions to either the assessment tool or the product.

Product-oriented evaluation

Product-oriented evaluation was characterized by Gros and Spector (1994) as a description and critical assessment of courseware made by an expert or experts from the same or different focus areas. This assessment was directed towards specific courseware. For the purpose of a product-oriented evaluation it was not necessary to use the courseware in an authentic context in order to judge the courseware's special features.

Gros and Spector (1994) indicated that several types of criteria have been proposed for product-oriented evaluations. These criteria included content analysis, the interactions of the courseware with the user, and the general usefulness of the program. Sorge et al. (1993) included other elements such as ease of initial learning, efficiency of use, ease of remembering interface items, error rates, and subjective responses to the system.

Observational evaluation

The purpose of observational evaluation was to collect data about the user's behaviour while they were using the courseware. The concept of observing the students as they worked through Internet-based courseware was introduced by Reiser and Kegelmann (1994) as a method of overcoming the problems associated with subjective evaluations. Reiser and Kegelmann suggested that evaluators could determine the quality of the courseware by observing the students as they worked through the Internet-based courseware. Through the observations the researcher could determine areas that the students were drawn toward, areas the students had difficulties navigating, and areas where the students' attention lacked focus. Harper (1988) and Owston and Wideman (1987) suggested that a good evaluation included observing actual use of the courseware in an authentic context.

Observations could take the form of direct observation, video recording, or verbal protocols where the user was encouraged to think aloud (Reiser & Kegelmann, 1994). The primary limitation to the observations was that it was an obtrusive evaluative measure and as such could affect user activity and performance (Aedo et al., 1996). A secondary concern for researchers was that the observation evaluation could be excessively time consuming.

Survey evaluation

Survey evaluation involved gathering data with the use of interviews or questionnaires for the purpose of gaining information on the user's opinions as well as their understandings of the courseware (Aedo et al., 1995). Northrup (1995) suggested gathering attitudinal data by asking specific questions about the instruction. Northrup suggested that the questions may include:

Did you understand how to navigate through the instruction?

Was the instruction confusing?

Were you able to make choices to view information that was relevant to you?

Were you able to navigate back to the screens you wanted to view again?

Lohr et al. (1995) used questionnaires before and after a specific courseware intervention for the purpose of establishing baseline data.

The primary limitation with interviews was that they tended to be excessively time consuming and were more manageable when used for small group data collection. Questionnaires facilitated data collection more efficiently than interviews when used with large groups (Aedo et al., 1995).

Criterion-based evaluation

Criterion-based instruments provided a good method of assessing the presence or absence of important measures. These measures tended to fall into 11 main categories as outlined in an extensive compilation by Wilkinson et al. (1997). These 11 categories were:

1. Site access and usability
2. Author identification
3. Information structure and design
4. Validity of content
5. Navigation within the document
6. Aesthetic and affective aspects
7. Resource identification and documentation
8. Authority of the author
9. Relevance and scope of content
10. Accuracy and balance of content
11. Quality of the links

These categories did not include measures for educational web sites. Additional educational categories were added by Symons (1997), which included such things as:

1. Learning objectives
2. Target audience
3. Social considerations
4. Instructional quality
5. Curriculum support
6. Appropriate pedagogy

There were, however, inherent problems with this type of evaluation. Rowland (1994) stated that criterion-based evaluations could be used to produce a checklist of criteria against which courseware could be compared. In spite of the creation of a checklist, criteria based on the assessment did not provide an overall assessment of the product. Rowland also indicated that since the criteria did not operate independent of each other, it was difficult to ensure that by changing certain criteria the worth of the courseware would increase. Another feature of the criterion-based evaluation was that it was typically executed by specialists or educators and not by the targeted users for which the curriculum was designed. Criterion-based assessment was an evaluation of the product itself and not the product's effect on the learner.

Typically criterion-based evaluations followed three methods of rating. The first was to examine the individual components, and sometimes weight the components differently to arrive at an overall rating. The second was to rate a feature based on a Likert-type scale indicating the degree to which each feature was present, while the final method was to simply identify if a feature was present or absent (Reiser & Kegelmann, 1994). This type of evaluation has also been called a product-oriented evaluation.

A combination of these approaches will capture a multidimensional judgment of the product's worth. The choice of a specific method depends largely on time constraints, the availability of users and experts, the type of data required, the stage of courseware development, and the cost of the evaluation method (Aedo et al., 1996). Reiser and Kegelmann (1994) conducted an assessment of different methods for evaluating for educational software. Their findings indicated that there was a need to involve the students as participants in the evaluation process. Reiser and Kegelmann also found that data should be collected for evaluating the effect of the software on the student.

Two different cohorts were established in the literature as valid adjudicators (Reiser & Kegelmann, 1994). These were the expert adjudicators and the end-user adjudicators.

Expert evaluation

Aedo et al. (1996) defined expert evaluation as a method that involved asking experts to judge the courseware and identify potential problems. The expert evaluation could originate from different orientations depending on who the expert was. For example, an educator may focus on the underlying pedagogy of the courseware, while a web site designer may be more likely to focus on principles of design (Mauldin, 1996). A benefit of expert evaluation was cost effectiveness. Expert evaluation could be accomplished with a small focus group of experts who detected significant problems (Aedo et al., 1996). Limitations associated with this type of evaluation included biased evaluations and the problems on inauthentic user behaviour (Aedo et al., 1996; Reiser and Kegelmann, 1994).

End-user evaluation

User-oriented evaluation focuses on the role of the learner as a participant in the evaluation process. User-oriented evaluation has been described by Gros and Spector (1994) as an evaluation method designed to assess the effects of the program on the user. User-oriented evaluation examines the interaction between the courseware and the user, and the users' responsiveness to the courseware.

Reiser and Kegelmann (1994) suggest that when subjective evaluations are used, students and teachers often rate the curriculum differently. The authors attribute this to the divergent features of programs that the different groups consider important. Smith and Keep (1986) generated opinions from 132 primary and secondary school children and found that excitement and audio-visual features were the primary evaluative criteria used by children to assess the instructional materials. The student had the unique opportunity to complete the curriculum as it was meant to be used, because they were the population for which the courseware was designed. This was an important aspect that external evaluators were not able to capture (Aedo et al., 1996).

Northrup (1995) indicated that a product evaluation completed by the target audience was an excellent method for gathering information of user perceptions of the instructional, technical, and style quality of the multi-media product. Sorge et al. (1993) suggested that an evaluator could get an accurate assessment of the programs worth only by trying it in an authentic context. End-users could also comment on such things as their motivation, interactivity, learner control, user interface, and screen layouts (Northrup, 1995).

A number of authors recommended using different methods of evaluation to capture the user-oriented perspective (Gros and Spector, 1994; Reiser and Kegelmann, 1994). This involved capturing the complete learning process and not just the knowledge obtained by the students due to the instructional intervention (Gros and Spector, 1994). Methods of evaluation that captured the relationship between the courseware and the learner are observations and performance appraisals.

Summary of the Literature Reviewed

The review of related literature included a discussion of the various uses of the Internet in education. These uses included communication, information retrieval, and information sharing. It was impossible to discuss the rationale behind evaluating educational web sites without first discussing the merits of the Internet in education.

Evaluation theory for Internet-based courseware revealed three main findings: a) formative evaluation by a multi-disciplinary team was deemed the optimal evaluation methodology for Internet-based courseware, b) summative evaluation was the practical type of evaluation that educators could use to determine whether or not to use Internet-based courseware in their learning environments, and c) a multidimensional model that incorporates more than one instrument and orientation was found to be desirable.

Two types of evaluation methods were examined: a product-oriented evaluation, and an outcome-oriented evaluation. Product-oriented evaluations included observation evaluation, survey evaluation, and criterion-based evaluations, whereas the evaluation of the outcome of the product was determined by a performance appraisal. The literature discussed the individual merits and limitations

of each approach and determined the unique features of the data derived from each method.

Two different cohorts were established as valid adjudicators. These were the expert adjudicators and the end-user adjudicators. The experts brought to the evaluation a thorough comprehension of some aspect of the web site and the end-users offer the orientation of the population for which the courseware was designed.

CHAPTER THREE: METHODOLOGY AND PROCEDURES

Introduction

This project assessed evaluative measures for Internet-based courseware using two different methods of evaluation by two different cohorts. These methods were product-oriented evaluation, and outcome-oriented evaluation. For the purpose of the present study, the term evaluation referred to the systematic appraisal of the value of the Internet-based courseware (Stufflebeam and Shinkfield, 1985).

The first method of evaluation was a product-oriented evaluation, where the term product referred to the specific courseware. The term product-oriented evaluation referred to the evaluation of the courseware as a product (Gros and Spector, 1994). Product-oriented evaluations typically take the form of criterion-based checklists, where the evaluator is looking for the presence or absence of specific criteria. For the purpose of the present study an expert product-oriented evaluation was conducted with the use of the Simpson Educational Web Site Evaluation Tool, and an end-user product-oriented evaluation was conducted with the use of the End-user Educational Web Site Evaluation Tool.

The second method of evaluation was an outcome-oriented evaluation of the web site www.bodymatters.com. The pretest scores of the end-users were recorded prior to the intervention of the web site, and then the posttest scores were recorded. This performance appraisal determined whether or not the web site was effective in improving the end-users' performance.

Description of the Research Methodology

A performance appraisal in the form of a preexperimental pretest/posttest design by a single group was used to determine if the web site www.bodymatters.com was effective in improving the end-users' performance. The independent variable

was the intervention of the web site www.bodymatters.com, and the dependent variable was the end-user scores on the posttest.

Two criterion-based checklists were used to generate checklist data. The three experts provided responses to the Simpson Educational Web Site Evaluation Tool based on their different orientations. The resulting data were Likert Scale data that ranged from strongly disagree, disagree, undecided, agree, and strongly agree. The experts were allotted a comment section where they provided qualitative, open-ended comments to anything that the evaluation tool did not encompass.

The end-users responded to the End-user Educational Web Site Evaluation tool which resulted in binomial data (yes/ no) or undecided for each question. The end-users were also allotted a comments section where they provided open-ended qualitative responses.

Research Design

Due to the multidimensional nature of the research this chapter is presented in such a way that the pilot work precedes the introduction of the research method, as the pilot work guided the research methodology. Figure 1 represents the structure of the research.

Pilot Study 1

The product-oriented evaluation of the web site www.bodymatters.com.

A group of 23 Health Studies majors completed an expert, product-oriented evaluation of Internet-based courseware. The Health Studies majors had completed a fourth year course entitled Developing Health Education Web Sites where they learned the basic design principals for educational web sites. These experts used an evaluation checklist entitled the Simpson Educational Web Site Evaluation Tool

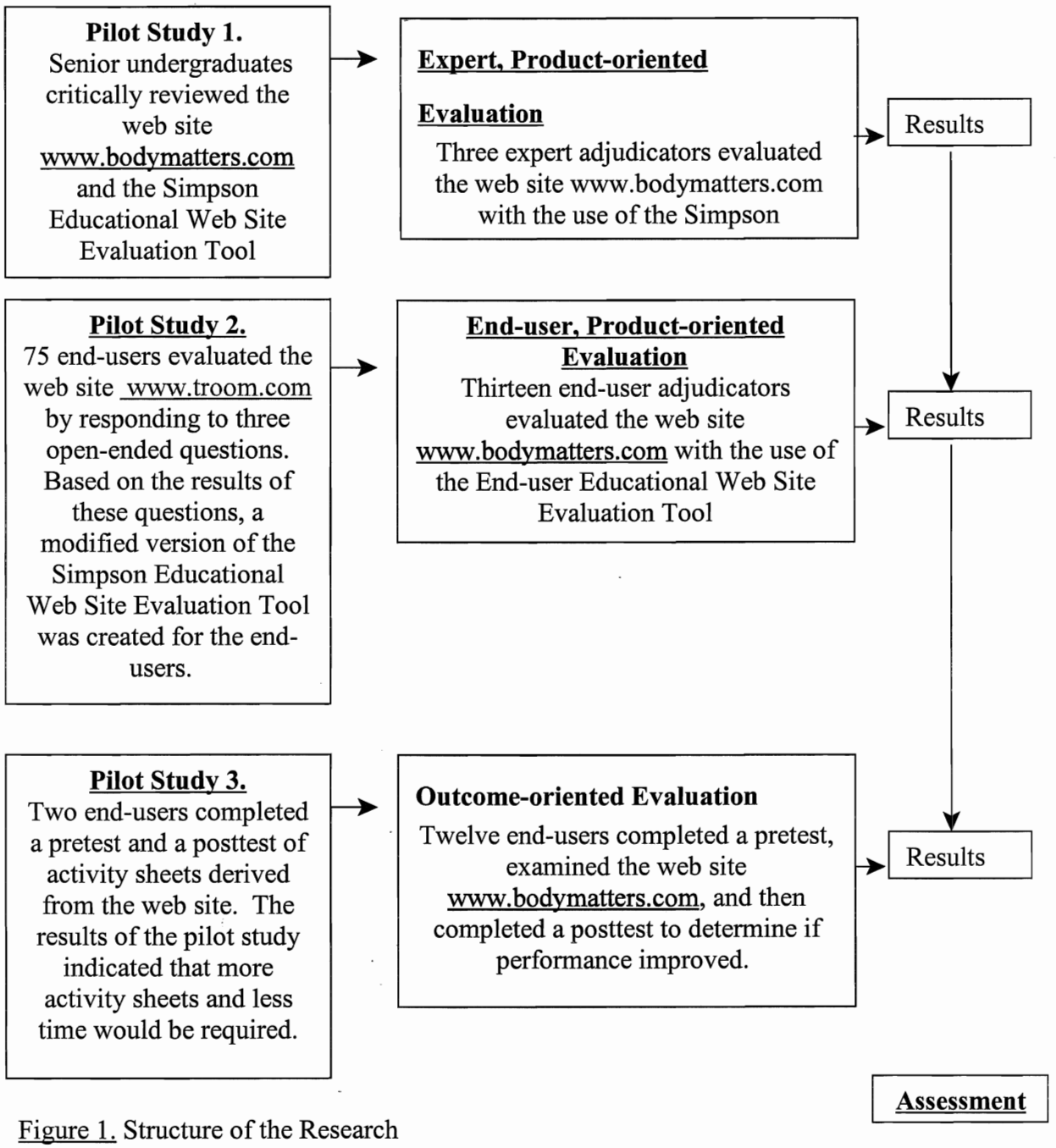


Figure 1. Structure of the Research

applied to the www.bodymatters.com web site. The experts were instructed to critically examine the www.bodymatters.com web site for the criteria outlined in the Simpson Educational Web Site Evaluation Tool.

Results of the Product-oriented evaluation of the web site www.bodymatters.com. The results of this product-oriented evaluation can be found in Appendix F. The 87 different criteria in the Simpson Educational Web Site Evaluation Tool were examined for either a positive or negative response by the expert adjudicators.

The responses of the expert adjudicators indicated that the web site www.bodymatters.com returned a good overall score of 80% for the Content Analysis section, with positive responses for the subsections as follows:

Defining the Web Site	78%
Credibility	71%
Social Considerations	83%
Teaching Considerations	91%
Total Score	80%

returned a moderate overall score of 68% for the Style Analysis section, with positive responses for the subsections as follows:

Design	71%
Features	58%
Navigation	78%
Teaching Considerations	66%
Total Score	68%

and returned an overall poor score of 47% for the Process Analysis section; with positive responses for the subsections as follows:

Input Considerations	42%
System Efficiency	80%
Teaching Considerations	40%
Support	27%
Total Score	47%

The poor results generated in the Process Analysis section of the web site www.bodymatters.com are likely a result of the Simpson Educational Web Site Evaluation Tool rather than the web site itself. A number of the criteria in the Process Analysis section were cited as questions that the students did not understand, which was also true for the Support section. Other sections, such as Input Considerations and Teaching Considerations, asked for criteria that web sites were not capable of, such as "Can the teacher add or change content?" or "Does the web site accept abbreviations?" These questions were based on the microcomputer tutorials and did not apply to educational web sites. Based on the pilot study results, these questions were removed and a new, shortened, 66-item checklist was created and employed as part of Evaluation Method 1.

Evaluation Method 1

Product-oriented evaluation conducted by expert adjudicators. The results of pilot study 1 revealed two things: a) the web site was a good example of Internet-based courseware, and b) the modified Simpson Educational Web Site Evaluation Tool was a good checklist to conduct a product-oriented evaluation. Based on these

preliminary findings, the researcher was confident in proceeding to the first level of assessment: the expert product-oriented evaluation. The experts who were selected to participate as subjects included a health educator, a classroom teacher, and a computer scientist. These experts were chosen to provide expertise in the processes of the web site, the style of the web site, and the content on the web site.

The researcher met with each of the three experts independently and explained the protocol for the completion of the Simpson Educational Web Site Evaluation Tool (Appendix G). The evaluator was instructed to indicate the degree to which they agreed with the criterion statement by placing a check mark in the appropriate box of the Likert scale. The evaluator was instructed to a) place a question mark in the undecided box if they did not understand the question, b) place an asterisk beside the criteria that they felt they had competency in, and c) leave the boxes blank for the criteria that the evaluator did not feel they had competency in. The experts were also advised to fill in the Comments Section with an overall impression of the web site, as well as any information that they felt the Simpson Educational Web Site Evaluation Tool did not capture.

These experts were then directed to the address of the web site www.bodymatters.com. The experts were advised that they could take as much time as required to concurrently examine the web site and complete the Simpson Educational Web Site Evaluation Tool.

Pilot Study 2

End-user product-oriented evaluation of the web site www.troom.com. A pilot study was conducted at the INVENTA conference in Welland Ontario, during

October 1997. Three groups of approximately 25 grade 7 students from the Welland County Separate School Board attended a 55-minute workshop designed to introduce students to evaluating web sites. A brief discussion about the evaluation process was conducted before the students were directed to work through the web site www.troom.com. The students were asked to respond to the following questions: a) Name 10 things that you like about the web site. b) Name 5 things that you do not like about the web site, c) List what you would do to improve the web site (Appendix H). The students were advised that they could exceed the number of responses requested.

Results of the end-user product-oriented evaluation. The results of these responses were grouped into the three categories of Content, Process, and Style. A content analysis revealed that a number of students thought the web site was a "good source of information, was educational, and helped people with questions that they were embarrassed to ask". Conversely, there were many comments that "the web site contained too little information for boys, was too personal and embarrassing, and was too commercial". The students cited the "Games" and the "Funny fill-ins" as the pages that they liked the most, while there was no real agreement on specific pages that the learners did not like.

The students had few comments on the process of the web site. Five respondents indicated that "it was easy to get around," while two respondents indicated that "it was difficult to get around." A single positive process response indicated that "the web site downloaded quickly." These results indicate that the

students did not reflect on the process elements of the web site.

The style analysis indicated that the students found "the web site had good graphics," was "colorful and funny," and "provided good variety." These results indicated that on the whole, the students liked the web site, however, there were a few respondents who indicated that "the pictures were too graphic," that "the room was a mess," and that "the web site was boring at times."

The responses to the question "What would you do to improve the web site?" included:

Content: Make it for both genders.

Use better music.

Add more games.

Add more Canadian content.

Style: Cut back on the graphic pictures.

Needs a warning screen to let boys know they have to look at tampons.

Provide a picture of Tina so people know who's advice they are asking for.

Have a chatroom.

Have a comment page.

Process: Make it easier to get around.

It is likely that grade 7 students were unable to think critically about the different process aspects of a web site. Although the task was broad, the students still provided many of the same comments as the expert, product-oriented evaluation. The responses indicated that many of the male students were uncomfortable with the

puberty-based content of the web site and focused most of their evaluation on the elements of the content that they disliked. In order to avoid these complications in the future, the researcher modified the Simpson Educational Web Site Evaluation Tool. The End-user Educational Web Site Evaluation Tool was created based on the same content, process and style categories of the Simpson Educational Web Site Evaluation Tool, and the language and concepts garnered from the student responses to what they liked and disliked about the web site. This ensured that the language was age appropriate and concept appropriate to grade 7 students. Also, as noted, the students did not speak to the process functions of the web site. To elicit these data, an end-user criterion-based checklist, which included criteria based on processor functions, was provided.

Evaluation Method 2

End-user product-oriented evaluation of the web site www.bodymatters.com. The second method of evaluation elicited the opinion of the participant. This evaluation was based on an end-user evaluation of the Internet-based courseware. Based on the findings in pilot studies 2 and 3 the researcher organized a session for the selected grade 7 students from Greendale Public School. The students completed a pretest to ascertain their baseline performance (Appendix J), and then they were given 45 minutes to examine the web site www.bodymatters.com before they completed a posttest (Appendix J).

The students were advised that after they completed the posttest, their opinion on the web site would be solicited. The students were provided with a copy of the End-user Educational Web Site Evaluation Tool (Appendix K). This tool was the

modified version of the Simpson Educational Web Site Evaluation Tool. The researcher walked the students through the first question to ensure that they understood the activity, and asked if there were any questions. The students' attention was then directed towards the comment sheet and they were instructed to write down any comments that they had about the web site. They were given 20 minutes to complete the evaluation. For the verbal protocol for the end-user intervention, please see Appendix L.

Pilot Study 3

End-user performance appraisal. A convenience sample of 2 students from the target age group completed a pretest that consisted of the activity sheets from the web site. They were then instructed to take some time to examine the web site www.bodymatters.com, and then to complete the posttest. The purpose of this pilot study was to determine the number of activity sheets the students could complete, as well as to determine the amount of time that the students should be given to examine the web site.

The results of the pilot study indicated that the students were not provided with enough activity sheets to make the use of the web site challenging. Based on these findings, the researcher increased the number of activity sheets that made up the pre and posttests from three activity sheets to six activity sheets. The time that the students were given to use the web site and complete the activity sheets was increased from 30 minutes to 45 minutes.

Evaluation Method 3

End-user performance appraisal. The third method of evaluation was experimental in the form of a performance appraisal. A performance appraisal was

conducted to evaluate the effects of the program on the learner (Stufflebeam & Shinkfield, 1985). For the purpose of this study the variable that was examined was the performance of the student after the intervention of the Internet-based courseware. This intervention was conducted with twelve13- year-old students from Greendale Public School in the Education computer lab at Brock University.

The end-users were asked to complete the activity sheets labelled #1 without the help of the web site or their fellow classmates (Appendix J). They were instructed that some of the questions did not have correct or incorrect answers and that their responses for these questions would not be marked. The end-users were advised to ask questions if they had any and that they would have 15 minutes to complete the task. The students completed the pretest in approximately 12 minutes.

The end-users were then advised that they would have 45 minutes to examine the web site www.bodymatters.com. They were instructed that this was a test of the web site and not of their ability. They were advised to avoid the "Answers" page of the web site because the results of any student who looked at the answer page would not be included in the study. They were also advised that if they had any questions about the content of the web site they could ask the researcher, or they could write their query on a piece of paper which would be passed on for their Public School Nurse to address. Finally, the students were instructed that if they found the information upsetting they could leave at any time.

All of the computers had been logged on to the web site www.bodymatters.com prior to the arrival of the students; however, the monitors were all turned off so the

students did not have access to the web site. At this point the students were instructed to turn on the monitors in front of them and to examine the web site www.bodymatters.com for 45 minutes.

Finally, after the 45 minutes, the students completed the posttest, which was the same set of activity sheets that they had completed for the pretest (Appendix J). The students were not given a time limit in which they had to complete the posttest; however, they were all finished within 10 minutes.

Selection of Participants

The Expert Adjudicators

The expert adjudicators were derived from a convenience sample of experts in the Niagara region. The experts provided insight into pedagogical considerations, web site design considerations, and content considerations for the web site www.bodymatters.com.

These experts provided the following responses to questions soliciting their background to ensure expert status (Appendix M).

Computer Scientist. The computer scientist completed a B.Sc. (Hons) in Computer Science from Brock University and is currently pursuing a Master of Science in Computer Science at Guelph University. He completed a Systems Administrator course and an Advanced Systems Administrator course from SGI and was employed as a Systems Administrator in the Faculty of Education at Brock University. This expert identified himself as having an advanced level of expertise in navigating the Internet. The computer scientist spent 2.5 hours completing the Simpson Educational Web Site Evaluation Tool

Health Educator. The Health Educator completed a B.A. in Psychology and was a Registered Nurse. She spent ten years as a Public Health Nurse, and 9 of those 10 years were spent teaching sexuality to students in grades 5 through OAC. This expert identified herself as having an intermediate level of expertise in navigating the Internet. The health educator spent 5.5 hours completing the Simpson Educational Web Site Evaluation Tool.

Teacher. The teacher completed a B.Sc. in Education, an M..Sc. in Education and 5 courses towards a Ph.D. He was also a Naval Flight Officer in the U. S. Navy where he worked with airborne and ground-based computers. This educator had the unique experience of conducting extensive beta-testing with educational software and attended computers-in-education workshops and conferences.

The teacher was employed as an aviator, a flight instructor, an outdoor pursuit teacher, a university lecturer, and an elementary school teacher (including sexual education). This expert identified himself as having an advanced level of expertise in navigating the Internet. This teacher spent 1.5 hours completing the Simpson Educational Web Site Evaluation Tool.

The End-user Adjudicators

The end-user adjudicators consisted of a convenience sample of 12 students from a grade 7/8 class from Greendale Public School located in Niagara Falls, Ontario, Canada. The students were five 13 year-old boys, six 13 year-old girls, and one 12 year-old girl. Three of the girls indicated that they had the Internet at home, and they indicated that they used it for less than an hour, 2 hours, and 3 hours respectively per week. One of the boys indicated that he had the Internet at home and

that he used it for approximately one hour per week. Four of the males and four of the females indicated that they did not have the Internet at home; however, one of the girls indicated that she used the Internet for approximately an hour a week at her cousin's house. The Health Educator taught these students their sexuality courses and indicated that the students had taken their grade 7/8 puberty education learning modules with her. The Teacher was the classroom teacher for these students. He indicated that there is a computer terminal in the classroom with different web sites loaded to the hard drive, and that the students have the opportunity to use the Internet in his classroom, as well as in their school library.

Instrumentation

The evaluation of the Simpson Educational Web Site Evaluation Tool. A pilot study was conducted to establish a measure of validity and reliability for the Original Simpson Educational Web Site Evaluation Tool (Appendix A). The purpose of the pilot study was to invite a critical review of the Simpson Educational Web Site Evaluation Tool, as well as an expert, product-oriented evaluation of the web site www.bodymatters.com. The Simpson Educational Web Site Evaluation Tool was developed from existing evaluative tools that were used to assess educational software and web sites (Alexander & Tate, 1996; Doll, 1987; Harris, 1997; Kirk, 1996; McLachlan, 1996; Squires & McDougall, 1994; Symons, 1997; Wilkinson, et al., 1997; Willing & Girard, 1990).

The Original Simpson Educational Web Site Evaluation Tool was evaluated by 23 senior undergraduate Health Studies majors enrolled in a course called "Developing Health Education Web Sites". These students used the Simpson

Educational Web Site Evaluation Tool to conduct a product-oriented evaluation of the "Body Matters" section of the www.tampax.com web site. Subsequently, the students answered an open-ended questionnaire that invited input on clarity and relevance of questions, comprehensiveness of the tool, and ease of use (Appendix B). The student adjudicators were informed that the evaluation tool was designed for educators to evaluate Internet-based curriculum.

Results of the evaluation of the Simpson Educational Web Site Evaluation Tool. The results (Appendix C) indicate that there were a number of criteria that the student adjudicators found to be unclear or irrelevant. Thirteen of the 23 adjudicators indicated that, if they had the opportunity, they would add a question to the evaluation tool. Seventeen of the 23 respondents indicated that the evaluation tool was easy to use, citing that "it was comprehensive and well laid out." The 6 respondents who indicated that the tool was difficult to use cited that "it was too long and tedious," and that "some of the questions were unclear." The Simpson Educational Web Site Evaluation Tool has been modified based on these results (Appendix D). The number of criteria was reduced from 87 to 66 in an effort to maintain the comprehensiveness of the tool, while limiting the length. The criteria that were identified as unclear were modified to clarify the language, and a definition sheet was included (Appendix E). Twenty of the 23 adjudicators responded positively to the question, "Do you think that this tool provides a valid evaluation of a web site?", two adjudicators were undecided, and one adjudicator responded negatively. Suggested improvements that were incorporated in the modified tool include: a) implementing a Likert rating scale, and b) soliciting input from the

learners. The pilot study was used to ascertain validity and reliability of the Simpson Educational Web Site Evaluation Tool.

The Creation of the End-user Educational Web Site Evaluation Tool.

The End-user Educational Web Site Evaluation tool was modified from the Simpson Educational Web Site Evaluation Tool. The End-user Educational Web Site Evaluation Tool included the same three categories of content, process, and style as the Simpson Educational Web Site Evaluation Tool; however, where the expert evaluation tool consisted of 66 criteria, the end-user evaluation tool consisted of only 20 criteria. The Likert scale of the expert evaluation tool was eliminated and replaced with binomial yes/no responses, and the criteria were listed as questions instead of statements. Finally, the language was modified to be appropriate for the end-users. The results of pilot study 2 helped to develop the wording for each question as the concepts and the language were derived from the open-ended responses of the end-users in pilot study two.

The Pretest and Posttest

The pretest and posttest used to ascertain the performance appraisals were derived from the activity sheets from the web site www.bodymatters.com. These activity sheets were used because they ensured that the information for the questions was found in the web site.

Data Collection and Recording

The expert responses to the Simpson Educational Web Site Tool were recorded directly on the tool itself. The comment section was attached to the evaluation tool and the experts wrote out their comments.

The end-user responses to the End-user Educational Web Site Evaluation Tool were recorded directly on the tool itself. The comment section was attached to the evaluation tool and the end-users wrote out their comments.

The performance appraisal consisted of a pretest and a posttest. The students completed a package that included one pretest, one posttest, one End-user Educational Web Site Evaluation Tool, and one open-ended comment sheet. The pretest, posttest and comment sheet in each section were numbered with the same participant identification number to ensure that they could be matched for data analysis. On the pretest the end-users indicated their age, their sex, and the number of hours per week that they spend using the Internet.

Data Processing and Analysis

A student t test determined whether or not the difference between the mean end-user scores on the pretest and the mean end-user scores on the posttest were a real difference rather than a chance difference. A t test was also used to determine whether or not the difference between the mean scores of the expert product-oriented evaluation of the web site www.bodymatters.com was statistically different than the end-user product-oriented evaluation of the web site www.bodymatters.com.

The comments section was subject to a content analysis based on the same content, process, and style considerations outlined by the Simpson Educational Web Site Evaluation Tool.

Likert scale data were generated from the expert product-oriented evaluation of the web site www.bodymatters.com. The experts provided Likert scale data that were based on a 5-point scale designed to measure their judgment about the different

criteria. The continuum of responses ranged from strongly disagree, disagree, undecided, agree, to strongly agree. The respondents were asked to place a check-mark on the response that best represented their judgment. A limitation of checklist data was that the items may not be weighted equally and therefore a statistical analysis of the Likert scale data was not possible. Statistical analysis was also undesirable due to the small population that was used to complete the checklists. Therefore the different criteria were examined individually and the total score of each section was examined.

Binomial data were generated from the end-user product-oriented evaluation of the web site. The end-users also had the option of responding undecided if they did not choose yes or no. The criteria from these responses were examined on an individual basis, and were compared with the expert product-oriented evaluation.

Methodological Assumptions

Performance Appraisal

If an individual performs poorly it will be assumed that it is the shortcoming of the intervention and not of the individual.

If an individual performs poorly it will be assumed that it is the shortcoming of the intervention and not processor-related problems such as unfamiliarity with the Internet.

The World Wide Web provided a viable learning experience.

Checklists

The Simpson Educational Web Site Evaluation Tool and the End-user Educational Web Site Evaluation Tool were comprehensive and valid evaluation tools.

Experts

The experts are qualified experts.

The Web Site

The web site www.bodymatters.com was specifically designed for grade 7. students.

Limitations

Limitations of the End-user Performance Appraisal

The limitations of the pretest/posttest design have been identified as testing, history, maturation, statistical regression, and instrumentation (Cook & Campbell, 1979). The threats to validity in this study included testing issues and sample selection limitations.

Testing was a threat to the validity of the findings because the pretest could have been the impetus to learning the correct answers to the items, thus increasing the posttest performance.

The population that was used in the current research was not randomly selected and cannot be considered representative of the larger population. This sample was also very small which further limits the generalizability of the study; however, the purpose of the current research was to provide a descriptive assessment of different summative evaluation methods that could be applied to Internet-based courseware.

History did not threaten the results of this study as there was no time delay between the pretest and the posttest to allow the influence of external events to affect the results. For this same reason maturation was not a limiting factor as the students had no time to become more experienced with the subject matter outside of the

intervention of the web site. Statistical regression did not threaten the validity of the findings because the whole group of 12 students completed both the pretest and the posttest. Finally, problems with instrumentation were not a factor in this experiment as the definition of the outcome measures were never changed, and the pretest and posttest were both marked by the same examiner to ensure consistency.

In the present study the division of participants by gender was equal; however, as noted by Northrup (1995), the results were limited by low racial and cultural diversity amongst the subjects. This factor must be considered when examining the results of the performance appraisal.

Limitations of Criterion-based Checklists

Rowland (1994) determined that checklists did not provide a valid assessment of the product as a whole. Criterion-based checklists are a very structured form of questionnaire. The primary limitations of a checklist include the fact that the criteria and the responses are predetermined. The respondent's evaluation was limited by the criteria and responses provided by the author of the checklist. This limitation was especially true for the end-user evaluation tool, for although the end-users were provided with the opportunity to express their opinion, their opinion was restricted by a checklist designed by an adult.

The Hawthorn Effect may have affected the results of the end-user criterion-based evaluation. The students may have reported different results because they were flattered to be chosen for the experiment, they were out of their regular school environment, they were in a university computer lab, and they may have thought that it would please the researcher to provide a favourable report.

Restatement of the Problem

The problem is two-fold: a) The many uses of the Internet in education are well established in the related literature, yet there is no established standardized method for evaluating Internet-based courseware; and b) educators do not participate in the development of Internet-based courseware, yet they are encouraged to use it in their classrooms. These problems create the need for standardized, summative evaluation methods for Internet-based courseware that can be implemented by the educator.

Summary of the Chapter

This project assessed evaluative measures for Internet-based courseware using two different methods of evaluation by two different cohorts. These methods were product-oriented evaluation and outcome-oriented evaluation. For the purpose of the present study, the term evaluation referred to the systematic appraisal of the value of the Internet-based courseware (Stufflebeam & Shinkfield, 1985).

The first method of evaluation was a product-oriented evaluation, where the term product referred to the specific courseware. The term product-oriented evaluation referred to the evaluation of the courseware as a product (Gros & Spector, 1994). Product-oriented evaluations typically take the form of criterion-based checklists, where the evaluator is looking for the presence or absence of specific criteria. For the purpose of the present study the expert product-oriented evaluation was conducted with the use of the Simpson Educational Web Site Evaluation Tool and the end-user product-oriented evaluation was conducted with the use of the End-user Educational Web Site Evaluation Tool.

The second method of evaluation was an outcome-oriented evaluation of the web site www.bodymatters.com. The pretest scores of the end-users were recorded prior to the intervention of the web site, and then the posttest scores were recorded. This performance appraisal determined whether or not the web site was effective in improving the end-user's performance.

CHAPTER FOUR: FINDINGS

Introduction

The purpose of this project was to assess evaluative measures for Internet-based courseware. Specifically, two entities were evaluated within the study: a) the outcome of the Internet-based courseware, and b) the Internet-based courseware itself.

The outcome of the product was measured by the results of a performance appraisal by a target cohort. This set of data was gathered from performance scores generated from the learning appraisal completed by the end-users, namely, the grade 7/8 students from a public school.

The evaluation of the product was completed by experts and end-users in independent sessions. Two sets of data were derived from the expert product-oriented evaluation: a) checklist data, and b) open-ended comments data. The data were based on the product-oriented evaluation of the www.bodymatters.com web site. The Simpson Educational Web Site Evaluation Tool was used to provide a set of Likert scale data for a content, process, and style analysis. The second set of expert data was open-ended survey data generated from the responses of the expert adjudicators to a comment section that followed the Simpson Educational Web Site Evaluation Tool.

Two sets of data were derived from the end-user's product-oriented evaluation: a) checklist data, and b) open-ended survey data. The checklist data were derived from the product-oriented evaluation of the www.bodymatters.com web site using a modified version of the Simpson Educational Web Site Evaluation Tool. Binomial data were gathered. The second set of end-user data was open-ended survey data generated from the responses of the end-users to an open-ended comment section that followed the End-user Educational Web Site Evaluation Tool.

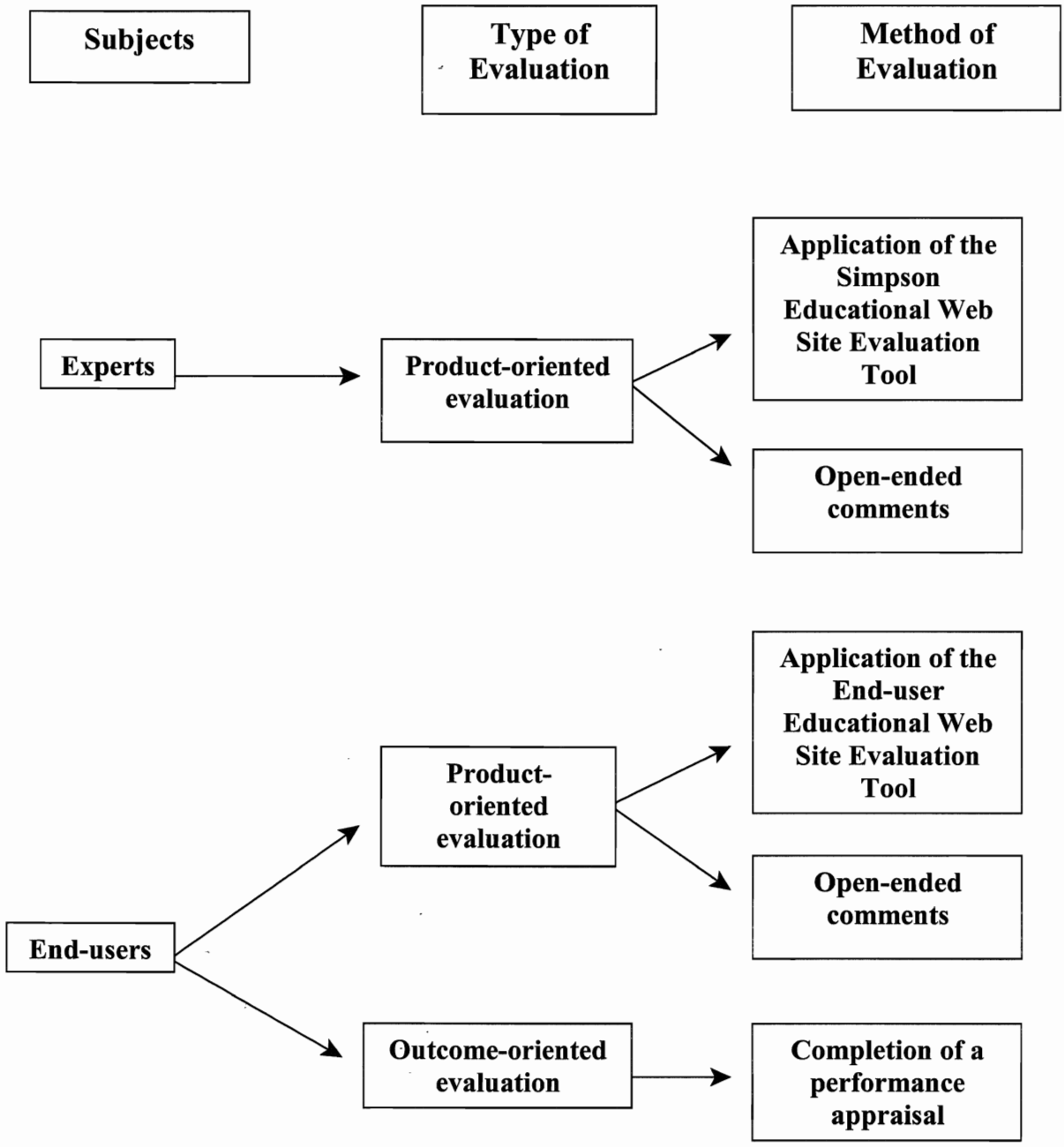
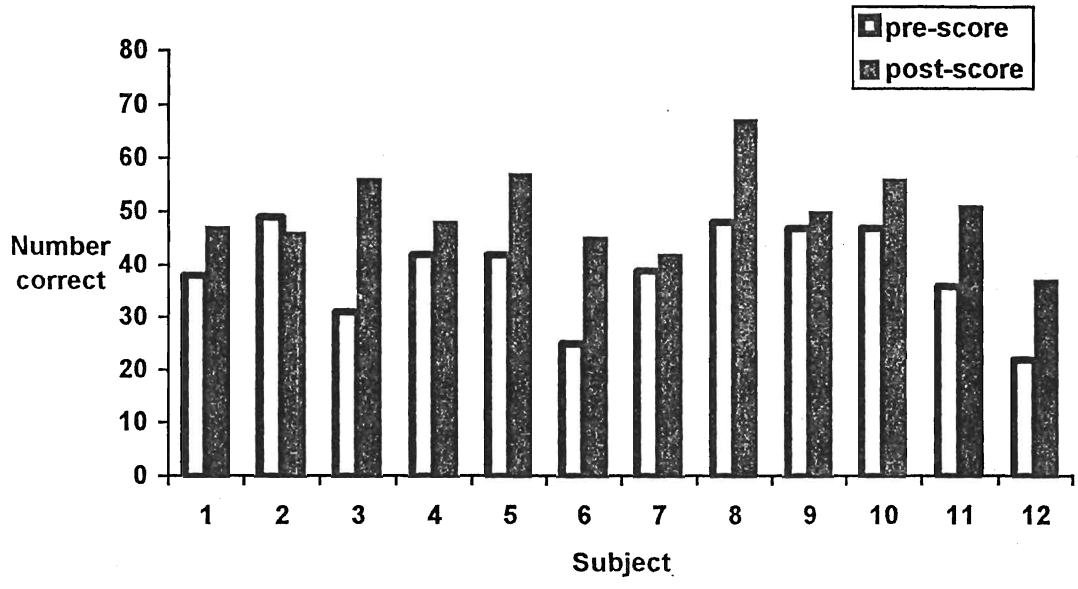


Figure 2. Structure of the Data Analysis



1.

Figure 3. End-user Performance Appraisal Results

The Findings

Results of the Evaluation of Learning Performance by the end-users

Figure 3 is a graphic representation of the student's performance on a one-group pretest-posttest design completed during a work session at Brock University. The pretest and the posttest were comprised of activity sheets (Appendix J) that were part of the web site www.bodymatters.com. The intervention was a 45 minute examination of the web site www.bodymatters.com by the students. Eleven of the 12 students performed better on the posttest than they did on the pretest. A t test was used to establish whether the improvement was statistically significant at 0.05. The null hypothesis: $x1 = x2$ was rejected because t observed was greater than t critical where t observed was 4.61 and t critical was 2.201. It was concluded that there was a positive, significant $p<0.05$ change in the students' performance based on the intervention of the web site www.bodymatters.com.

The Results of the expert, product-oriented evaluation of the web site www.bodymatters.com

The results of the expert, product-oriented evaluation indicated that there was no consensus among the experts for most of the criteria outlined by the Simpson Educational Web Site Evaluation Tool. The results were based on the application of the Simpson Educational Web Site Evaluation Tool to the web site www.bodymatters.com. The protocol for the completion of the Simpson Educational Web Site Evaluation Tool was as follows: The experts were advised to refrain from responding to criteria that they did not have competency in, to place a question mark beside criteria they did not understand, and to place an asterisk beside the criteria that they felt they had competence in (Appendix G). This third criterion created a

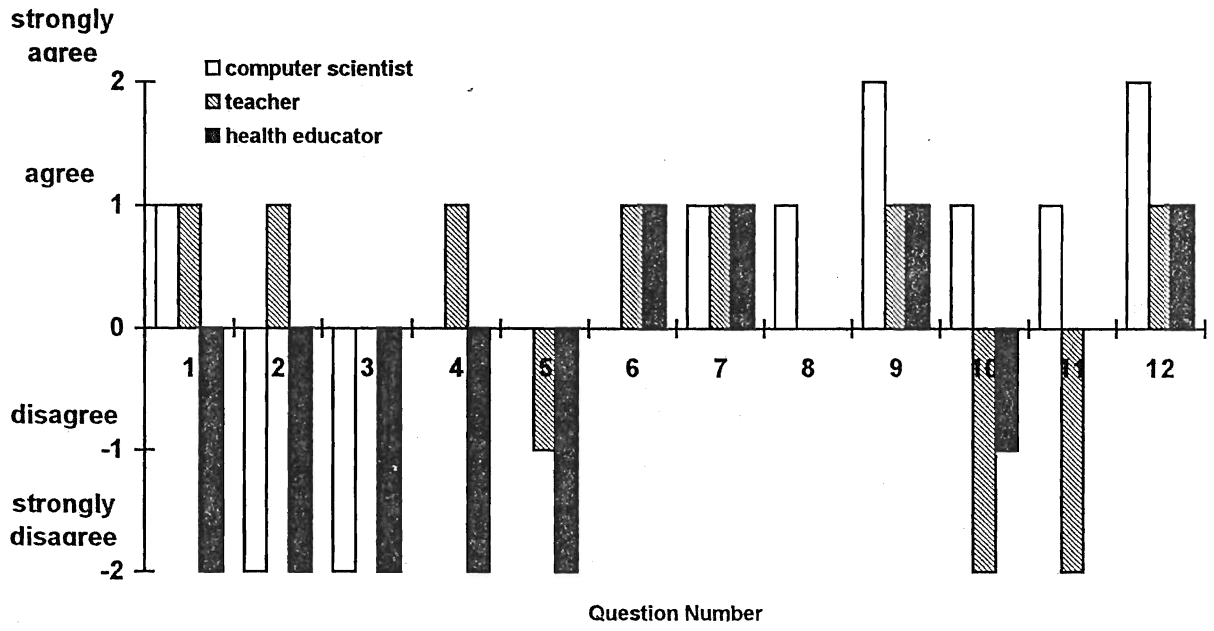


Figure 4. Results of the Expert Content Analysis-Defining the Web Site

Defining the Web Site

1. The topic is clearly defined.

2. The purpose of the web site is stated.

3. The objectives of the web site are stated.

4. Adequate information is provided to meet the objectives.

5. The target audience is defined.

6. The information is accurate

7. The information is current.

8. The web site is updated frequently.

9. Spelling, grammar, and punctuation are correct.

10. The links are relevant and appropriate to the topic.

11. The rationale for providing the links is explicit.

12. The content stands alone.

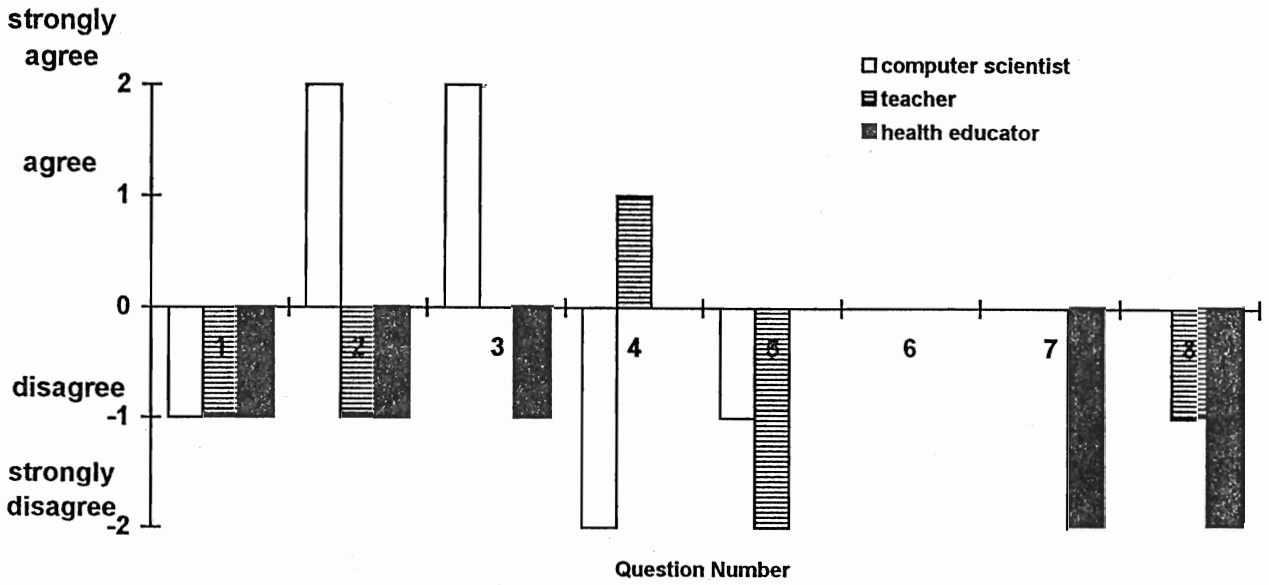


Figure 5. Results of the Expert Content Analysis- Credibility

Credibility

1. The identity of the author is provided.

2. Information about the author's training and education is provided.

3. The author's education and training are from respected organizations.

4. The author provides contact information beyond an e-mail address

5. References and a bibliography are provided.

6. The web site has been externally evaluated.

7. The company producing the web site provides unbiased information

8. All aspects of the topic are presented.

self-named subset of the expert population which will be referred to as the competent experts.

The results indicate that there was no consensus among the experts on the evaluation of the web site except for the social considerations section in the content analysis, and the system efficiency and teaching considerations section in the process analysis of the Simpson Educational Web Site Evaluation Tool.

Figure 4 is a graphic representation of the responses by the experts to the defining the web site section of the Simpson Educational Web Site Evaluation Tool. The responses for Figures 4 though 7 are based on a content analysis of the web site www.bodymatters.com.

The Teacher indicated competence (by placing a asterisks beside the criteria) in all of the criteria in the defining the web site section, while the Health Educator indicated competence for criteria 2 through 7, and the Computer Scientist indicated neither competence nor incompetence.

There were only two instances of consistent responses by experts who indicated competence on the selected criteria. The "competent" experts agreed that the information was accurate, and that the target audience was not defined. With the exception of these two criteria, the results were inconclusive.

Figure 5 is a graphic representation of the responses of the experts to the credibility section of the Simpson Educational Web Site Evaluation Tool. The Teacher indicated competence in all of the questions in the credibility section, while the Health Educator indicated neither competence nor incompetence for criteria 7 and 8, and the Computer Scientist indicated neither competence nor incompetence for all criteria.

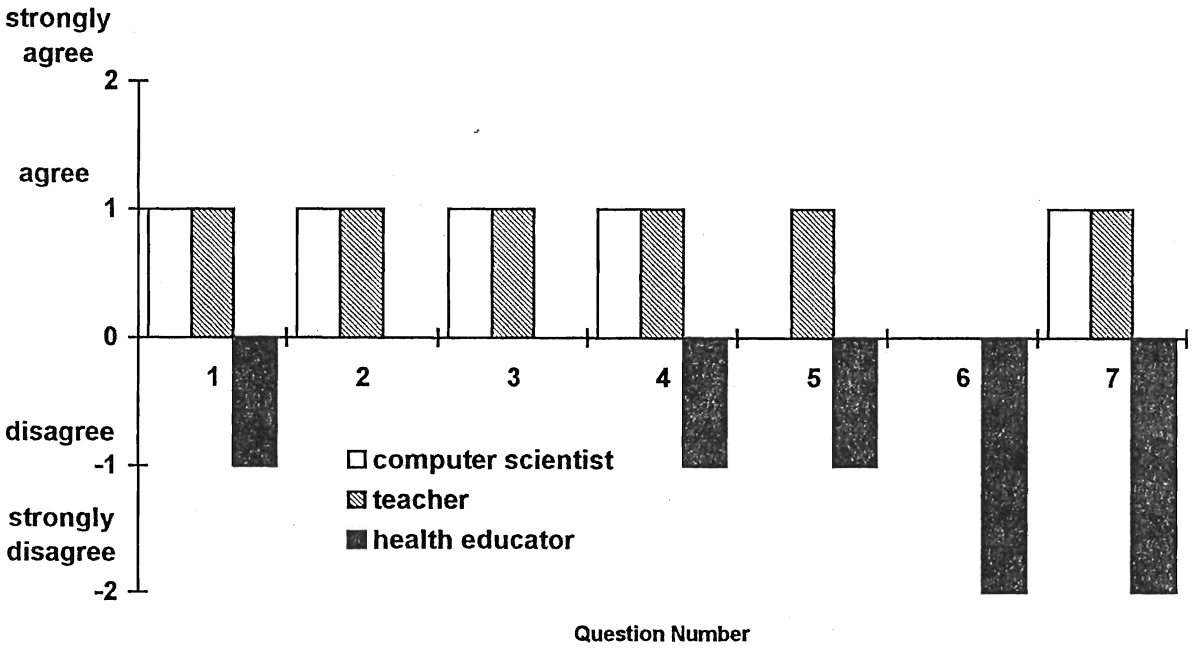


Figure 6. Results of the Expert Content Analysis - Social Considerations

Social Considerations

1. People of different races, ages, physical and mental abilities are featured in the web site.

2. Different groups are presented in positive ways.

3. The web site avoids an obvious cultural bias.

4. The social messages in the content are positive.

5. The web site promotes inclusion and acceptance.

6. The web site offers positive role models.

7. The web site helps the students develop decision-making skills.

The results of the credibility section indicated that there were only two instances of consistent responses by experts who indicated competence on the selected criteria. The educators indicated that the web site did not cover all aspects of the topic. All three experts indicated that the identification of the author was not provided. This criterion has been verified and the identity of the author was provided, therefore the experts provided an incorrect response.

The criterion statement, "The author provides contact information beyond an e-mail address" generated divergent responses. Upon verification of this criterion it was found that the web site www.bodymatters.com does provide a mailing address for the author. This criterion was the third criterion not answered correctly by the experts.

The Computer Scientist disagreed and the Teacher strongly disagreed with the criterion statement, "References and a bibliography are provided." The Health Educator indicated no competence and chose not to respond this criterion. A bibliography was provided for this web site; therefore this criterion was not answered correctly by the experts.

Figure 6 is a graphic representation of the responses of the experts to the social considerations section of the Simpson Educational Web Site Evaluation Tool. The results indicate a measure of consensus between the responses of the Computer Scientist and the responses of the Teacher. The Teacher indicated competence in all of the questions in the social considerations section, while the Computer Scientist and the Health Educator indicated neither competence nor incompetence.

The three experts did not agree on a single criterion within the social considerations section; however, the Teacher and the Computer Scientist agreed on five of seven criteria.

Figure 7 is a graphic representation of the responses of the experts to the teaching

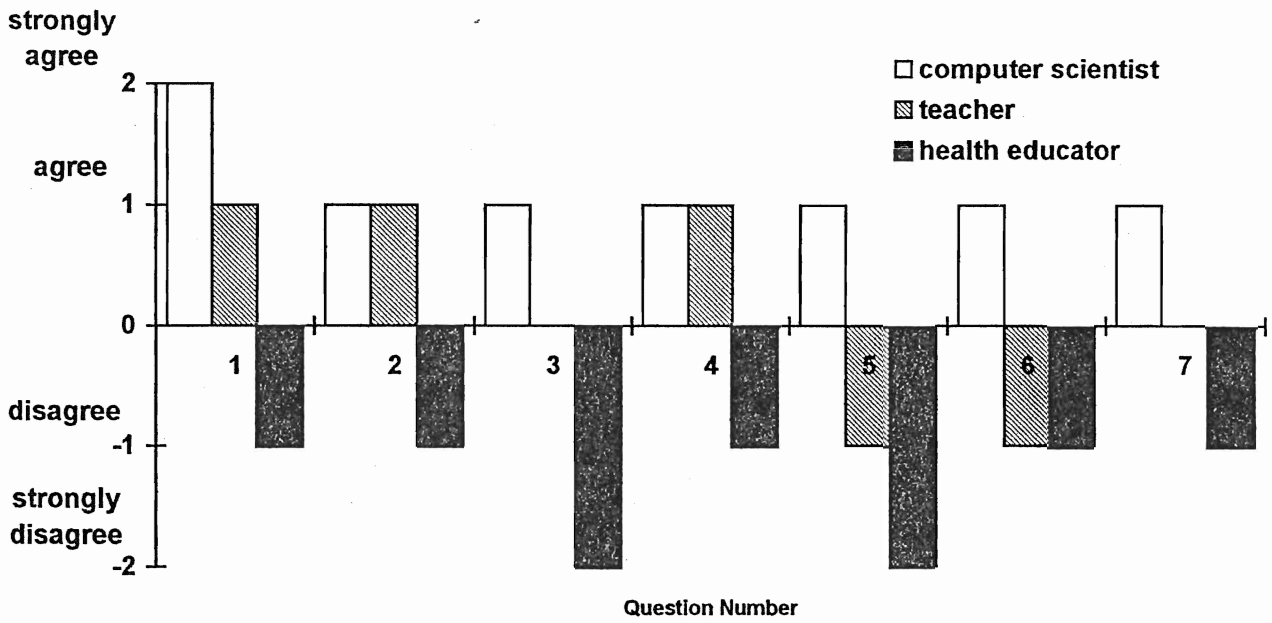


Figure 7. Results of the Expert Content Analysis - Teaching Considerations

Teaching Considerations

1. The focus of the web site is educational.
2. The web site supports the school curriculum.
3. The reading level is appropriate for the target audience.
4. The concepts are age-appropriate for the target audience.
5. Definitions are provided when necessary.
6. The content is engaging.
7. The content is comprehensive.

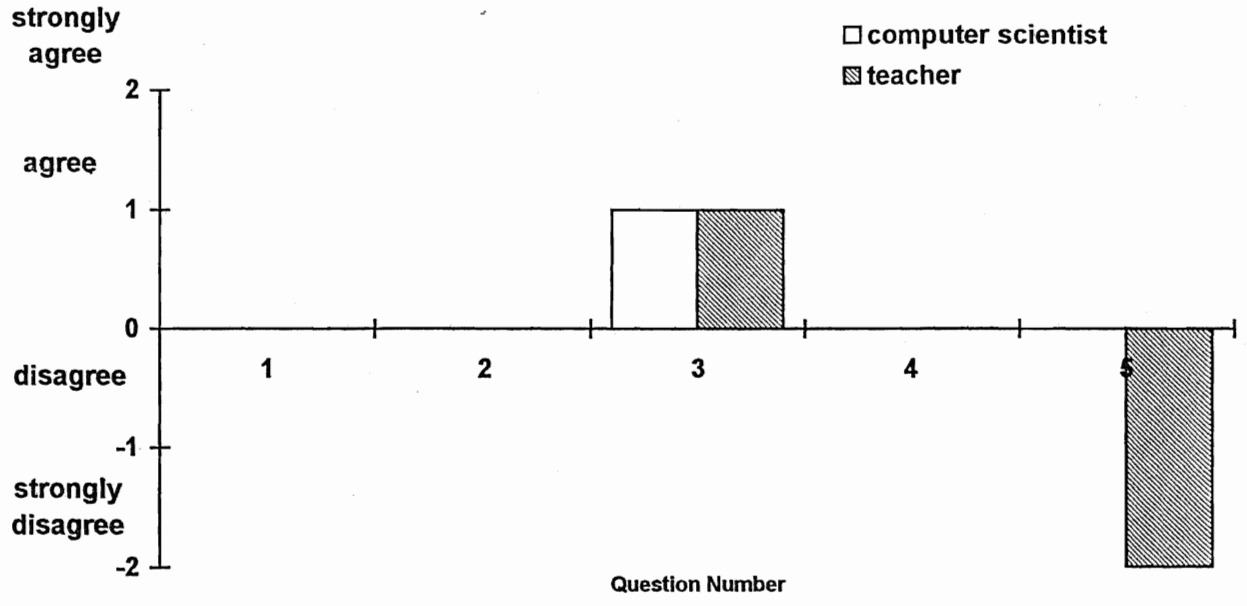


Figure 8. Results of the Expert Style Analysis - Design

Design

1. Graphics and animation are clear and easily interpreted.
2. Graphics and audio are appropriate to the population.
3. Characters are of legible size.
4. Sound is used with purpose.
5. The format is varied and interesting.

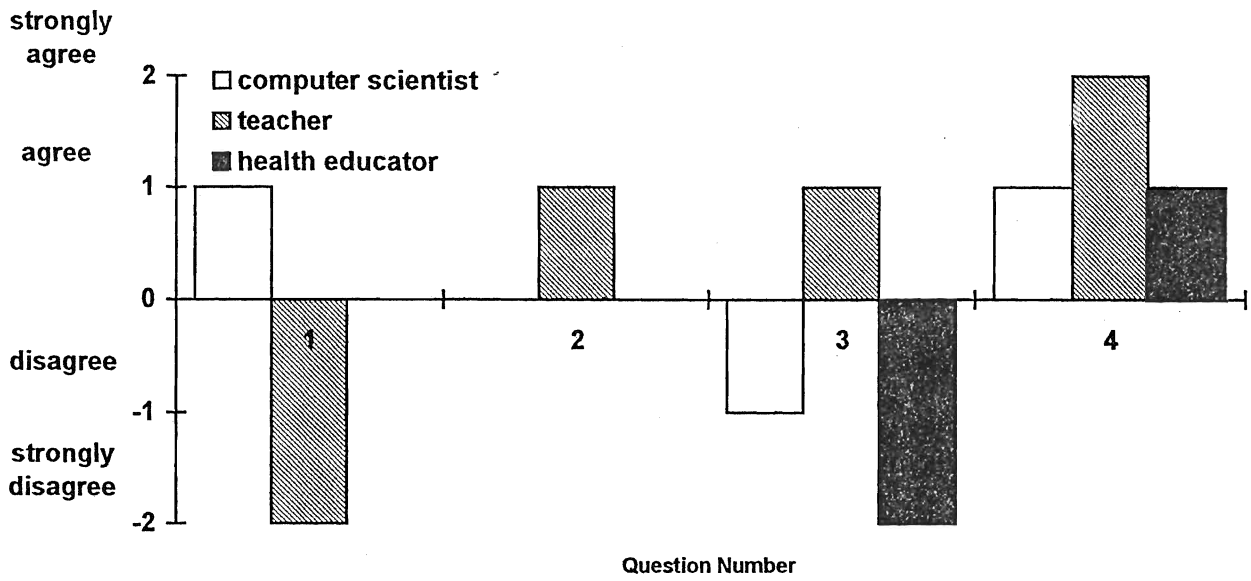


Figure 9. Results of the Expert Style Analysis - Features

Features

1. The web site uses interactive features.
2. If calculations are required, they can be done on screen.
3. Instructions are provided for all aspects of the web site.
4. Menus are provided to make the web site user friendly.

considerations section of the Simpson Educational Web Site Evaluation Tool. The Health Educator and the classroom Teacher indicated competence for the seven criteria in the teaching considerations category of the content analysis. The Computer Scientist indicated neither competence nor incompetence. The three experts did not respond consistently to any of the criterion statements; however, the Teacher and the Computer Scientist responded in similar directions on three of seven criteria. The Teacher and the Health educator were consistent on only two of the seven responses.

Figure 8 is a graphic representation of the responses of the experts to the design section of the Simpson Educational Web Site Evaluation Tool. The responses for Figures 8 through 11 were based on a style analysis of the web site www.bodymatters.com.

As the graph indicates, the experts did not respond to most of the criterion statements. The Health Educator indicated that she did not have competence in the design category by choosing not to respond to any of the criteria. The Teacher and the Computer Scientist indicated competence in all five criteria in the design section.

The Teacher indicated "not applicable" to the following criterion statements: "Graphics and animation are clear and easily interpreted," "Graphics and audio are appropriate to the population," and "Sound is used with purpose." The "not applicable" response to the criterion about sound was accurate because sound was not used in the web site. However, graphics were used in the web site. The Teacher did not indicate why the response to this criterion was "not applicable." The Computer Scientist did not respond to the same questions for which the Teacher indicated "not applicable."

Figure 9 is a graphic representation of the responses of the experts to the features section of the Simpson Educational Web Site Evaluation Tool. The Teacher and the Computer Scientist indicated competence in the four criteria of the features

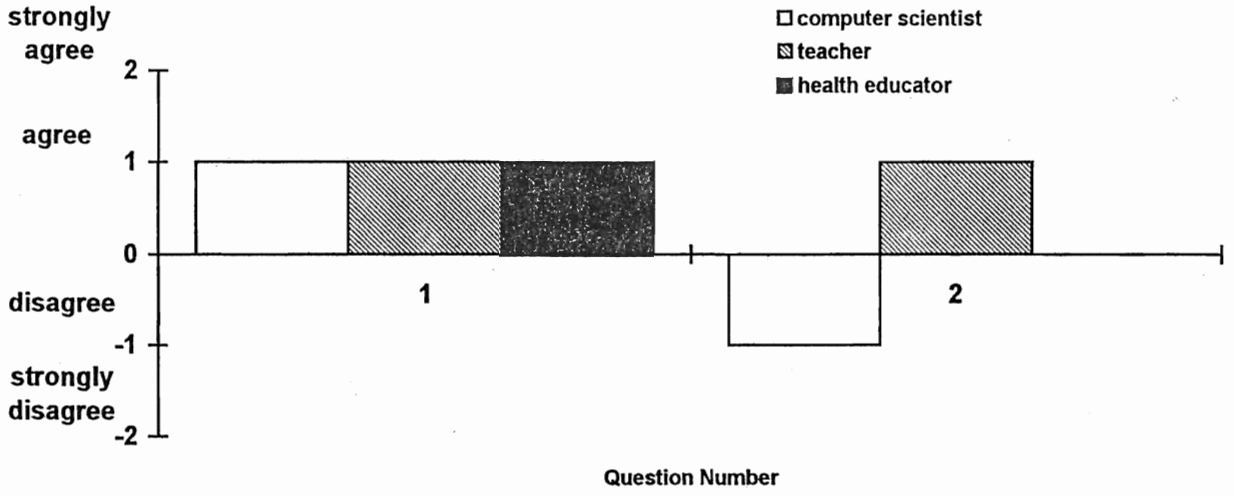


Figure 10. Results of the Expert Style Analysis - Navigation

Navigation

1. Each section is labelled with a heading.
2. Image maps are provided.

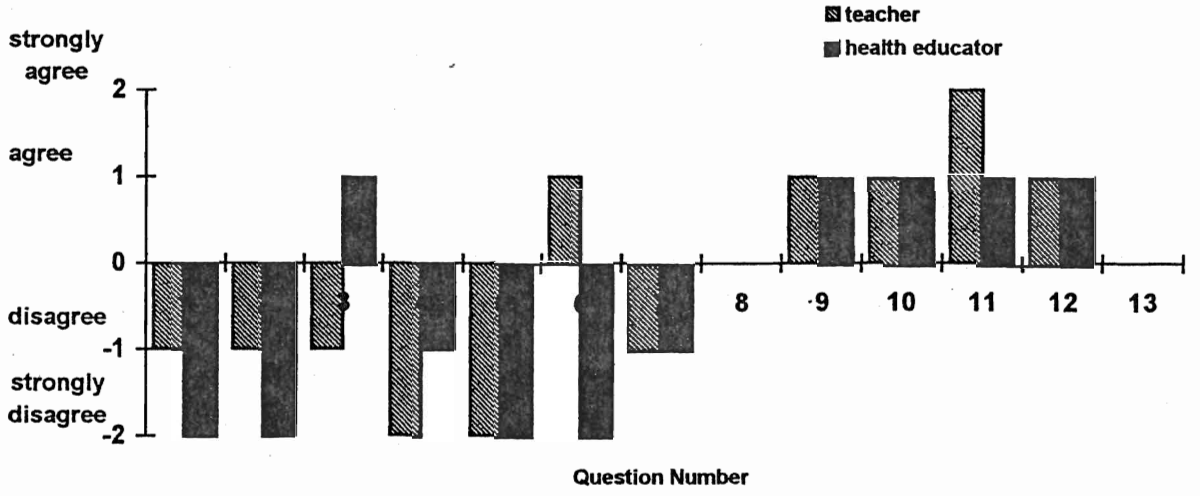


Figure 11. Results of the Expert Style Analysis - Teaching Considerations

Teaching Considerations

1. The pedagogic approach is superior using this medium.
2. The pedagogy is innovative.
3. Instructional strategies are appropriate.
4. Demonstrations are clear.
5. Minimal teacher supervision is required.
6. There are opportunities to answer open-ended questions.
7. Students can hypothesize and experiment.
8. The program provides a forum for student interaction.
9. The time limits of the class are addressed by providing workable units.
10. The material is organized and well sequenced.
11. The web site provides criteria to assess learning.
12. The questions in the test measure the mastery of the content.
13. Student responses match the program objectives.

category. The Health Educator indicated no competence for the first two criteria The Teacher agreed with the criterion statement, "If calculations are required, can they be done on screen". This is not an accurate response, as there is no method of doing calculations in the web site. The Computer Scientist did not respond to the criterion, although he indicated competence. The Health Educator indicated no competence and chose not to respond.

The Computer Scientist and the Health Educator agreed, and the Teacher strongly agreed with the criterion statement, "Menus are provided to make the web site user friendly." This is one of the rare criteria where the experts responded consistently.

Figure 10 is a graphic representation of the responses of the experts to the navigation section of the Simpson Educational Web Site Evaluation Tool. The Teacher and the Computer Scientist indicated competence for the navigation category and the Health Educator indicated neither competence nor incompetence.

Figure 11 is a graphic representation of the responses of the experts to the teaching considerations section of the Simpson Educational Web Site Evaluation Tool. The Computer Scientist indicated no competence in the teaching considerations section and chose not to respond. The Teacher indicated competence, and the Health Educator indicated neither competence nor incompetence.

The responses of the Teacher and the responses of the Health Educator were consistent. The experts provided identical responses for 5 of 11 criteria, and responded in the same direction (agree, strongly agree) for 9 of the 11 criteria. The educators were not in agreement with only two of the criterion statements. These

Table 1.
Overall Score of the Expert Application of the Simpson Educational Web Site Evaluation Tool.

Section	All expert scores	Total potential score	Competent expert scores	Total potential scores
Defining the web site	0	+/- 64	-6	+/- 34
Credibility	-10	+/- 36	-6	+/- 18
Social considerations	4	+/- 34	6	+/- 14
Teaching considerations	2	+/- 42	-4	+/- 28
Design	0	+/- 6	0	+/- 6
Features	2	+/- 18	3	+/- 16
Navigation	3	+/- 10	2	+/- 8
Teaching considerations	-7	+/- 46	-2	+/- 24
System efficiency	13	+/- 26	11	+/- 22
Teaching considerations	1	+/- 4	2	+/- 2

Note.

Calculation of All Experts Scores: This is the total score for all of the responses provided by the experts.

Calculation of the Competent Expert Scores: This is the total score for the self-named subset of experts who indicated with a astericks the criteria that they had expertise in.

Calculation of the potential scores: This score is the maximum and minimum score possible for each section of the Simpson Educational Web Site Evaluation Tool. If, for example, all three experts answered one question, the calculation would be represented by 3 experts X a potential score for the criteria of -2(strongly disagree) / +2(strongly agree) for a total potential score of -6/+6 for that criteria. The total potential scores of each criterion in a section were totalled to provide the Total Potential Score.

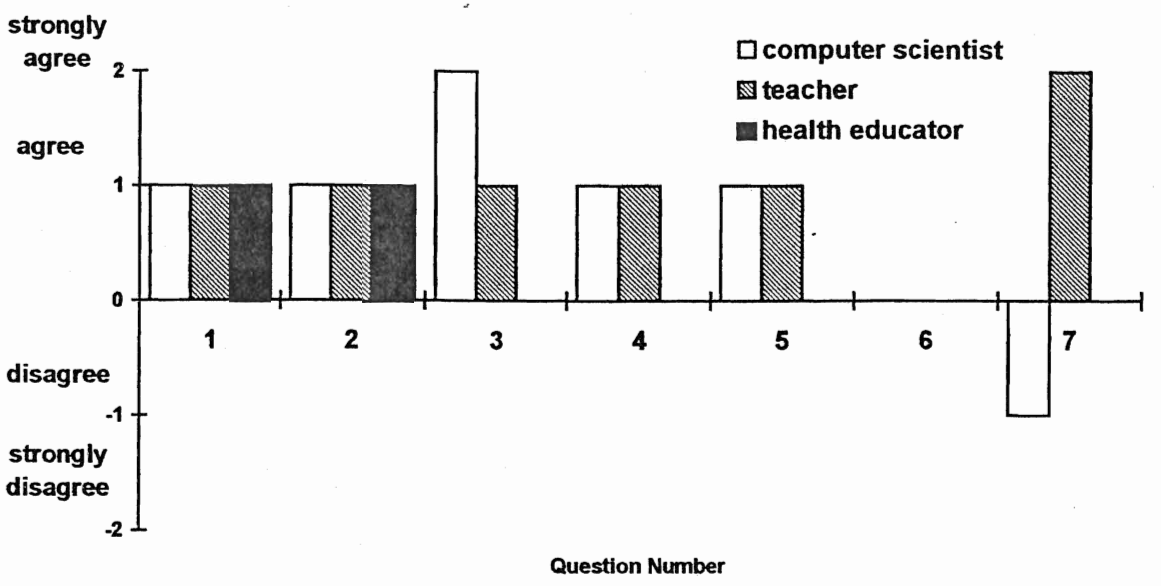


Figure 12. Results of the Experts' Process Analysis - System Efficiency

System Efficiency

1. The homepage loads smoothly and efficiently.
2. There is efficient transition between screens.
3. All of the pages download.
4. The web site can run without any special software requirements.
5. The text can stand alone if the graphics are not loaded.
6. Out-of-date links are quickly removed.

Teaching Considerations

7. Teachers are provided with user guides for the web site.

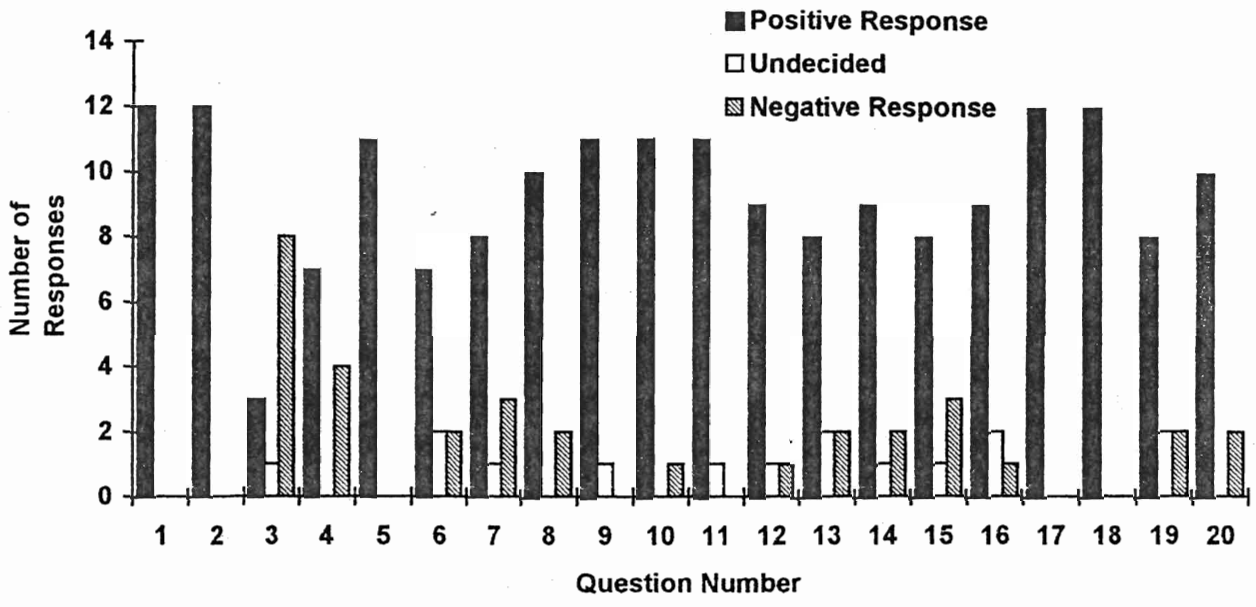


Figure 13. End-user Responses to the Criterion-based Checklist

Content Analysis

1. Is the web site a good source of information?

2. Has the web site taught you things you did not already know?

3. Does the web site have words you do not understand?

4. Does the web site address topics you do not understand?

5. Is this a good web site for people your age?

6. Are people from different cultures represented in the web site?

7. Do you have questions that are not answered by the web site?

Style Analysis

8. Is the web site interesting?

9. Are the graphics clear?

10. Are the graphics suitable for people your age?

11. Is the web site well laid out?

12. Does the web site offer choice and variety?

13. Are the activities enjoyable?

14. Does the web site have enough interactive features?

15. Is the web site boring?

16. Are instructions provided when you need them?

Process Analysis

17. Is it easy to move around the web site?

18. Do the pages of the web site load quickly?

19. Do all of the pages download?

20. Did you get lost in the web site?

statements were: "Instructional strategies are appropriate" and "There are opportunities to answer open-ended questions."

The educators did not respond to two criteria. These criteria were, "The program provides a forum for student interaction," and "Student responses match the program objectives." The Teacher indicated that the criterion statement "Student responses match the program objectives" is not applicable. The Teacher indicated that objectives were not provided for the web site, which would likely be why his response was that the criterion was not applicable. The Health Educator does not indicate why she did not respond to the criteria.

Figure 12 is a graphic representation of the responses of the experts to the system efficiency and teaching considerations sections of the Simpson Educational Web Site Evaluation Tool. The responses are based on a process analysis of the web site www.bodymatters.com.

The Teacher indicated competence in all of the criteria in the process section while the Computer Scientist indicated competence in the first five criteria. The Health Educator indicated neither competence nor incompetence for the first two criteria, and indicated incompetence for criteria 3 through 7. The expert's responses for this section were consistent.

Table 1 is a chart depicting the scores of each section of the Simpson Educational Web Site Evaluation Tool.

The End-user Responses to the Criterion-based Checklist

Figure 13 is a graphic illustration of the application of the End-user Educational Web Site Evaluation Tool (Appendix K) by 12 grade 7 students (age 13 years) during a work session at Brock University. The responses were scored as binary items (yes/no).

Yes was the positive response for the majority of the questions; however, to avoid confusing, false-negative questions, no was the positive response for five of the questions. The checklist was organized according to content, style, and process questions.

Content analysis. Questions1 through 7 in Figure 13 are included in the content analysis section. The results demonstrated that all of the students indicated that the web site was a good source of information, that it taught them things they did not already know, and that the web site www.bodymatters.com was good for people their age. Figure 13 also showed that 65% of students indicated that people from different cultures were represented in the web site. Forty percent of the students indicated that the web site addressed topics they did not understand, and that they had questions that were not answered by the web site. Finally, the majority of students indicated that the web site had words that they did not understand.

The results of the end-user content analysis indicated that the students found the web site to be informative, but that there were some areas where the comprehension of the material was hindered by inappropriate language and concepts they did not understand.

Style analysis. The results of the style analysis were primarily positive. The students indicated that the graphics were clear and suitable for people their age, that the web site was well laid out, and that instructions were provided when they needed them. There were some minor divergent responses to the indicators of the web site being interesting or boring. Three of 11 students indicated that the web site was boring, and 2 of the 12 students indicated that the web site was not interesting. This may be a result of 2 of the 11 students who indicated that there were not enough interactive features, 2 of the 10 students indicated that the activities were not enjoyable, and 1 of 10 students

indicated that the web site did not offer choice and variety.

Process analysis. Figure 13 indicated that all of the students found it easy to move around the web site, and that all of the pages of the web site load quickly. Most of the students responded that they did not get lost in the web site, and that all of the pages download.

A Comparison of the Results between the End-user Responses and the Expert Responses to the Criterion-based Checklists

Figures 14-16 are graphic representations of the responses of the experts and the responses of the end-users to two separate criterion-based checklists. The expert responses were based on criteria from the Simpson Educational Web Site Evaluation Tool. Their responses are in regular text. The end-users responses were based on a version of the Simpson Educational Web Site Evaluation Tool that has been modified to be language and age appropriate for grade 7/8 students. Their responses are in bold text.

The researcher matched the criteria provided to the end-users to the criteria provided to the experts. The end-user criterion-based checklist contained only 20 criteria where the expert criterion-based checklist contained 66 criteria. In order to match the criteria, the expert criteria were grouped into categories detailed by the 20 criteria for the end-user criterion-based checklist (e.g. question number 7). The positive percent scores were based on the number of positive responses compared to the total number of responses provided by the experts and the end-users. The end-users provided binomial responses (yes/no) where the experts provided Likert scale responses that ranged from strongly disagree to strongly agree. The expert responses that were either agree or strongly agree were considered positive responses. The end-user responses that were the positive response (either yes or no) were considered positive responses. An undecided response was counted as a response and was included in the total number of responses.

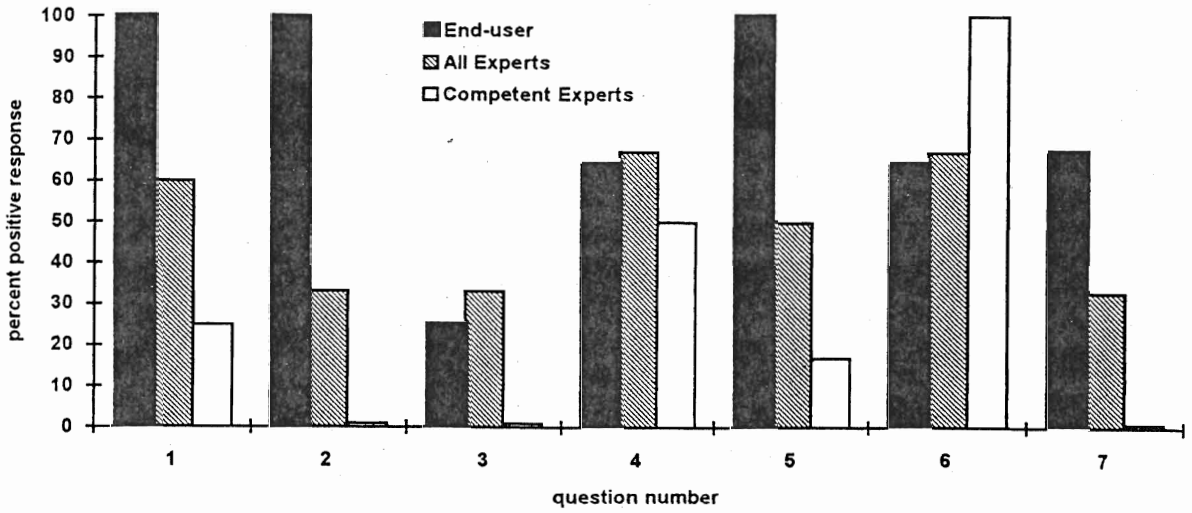


Figure 14. Comparison of the End-user and Expert Content Analysis

1. Is the web site a good source of information?
Adequate and relevant information is provided to meet the objectives
The focus of the web site is educational

2. Has the web site taught you things you did not already know?
The content is comprehensive.

3. Does the web site have words you do not understand?*
The reading level is appropriate for the target audience.

4. Does the web site address topics you do not understand?*
The concepts are age appropriate for the target audience.

5. Is this a good web site for people your age?
The reading level is appropriate for the target audience.
The concepts are age-appropriate for the target audience.
The content is engaging.

6. Are people from different cultures represented in the web site?
People of different races, ages, physical & mental abilities are included.

7. Do you have questions that are not answered by the web site?*
The content is comprehensive.

**** indicates that no is the positive response***

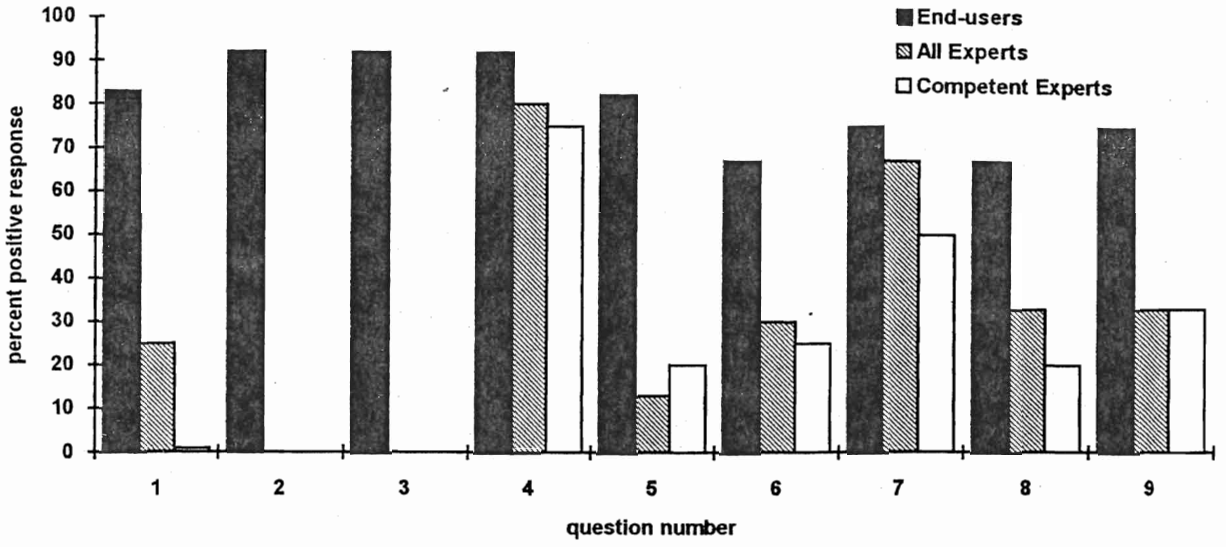


Figure 15. Comparison of the Expert and End-user Style Analysis

1. Is the web site interesting?

The style is engaging.
The format is varied and interesting

2. Are the graphics clear?

Graphics and animation are clear and easily interpreted.

3. Are the graphics suitable for people your age?

Graphics and audio are appropriate to the population.

4. Is the web site well laid out?

Characters are of legible size.
The format is varied and interesting.
The material is organized and well sequenced.

5. Does the web site offer choice and variety?

The format is varied and interesting.
There are opportunities to answer open-ended questions.
Students can hypothesize and experiment.
The style is engaging.

6. Are the activities enjoyable?

The format is varied and interesting.

The web site uses interactive features.
There are opportunities to answer open-ended questions.
Students can hypothesize and experiment.
The style is engaging.
The program provides a forum for student interaction

7. **Does the web site have enough interactive features?**
   The web site uses interactive features.
   If calculations are required, they can be done on screen.
   The program provides a forum for student interaction

8. **Is the web site boring?**
   The web site uses interactive features.
   The format is varied and interesting.
   The style is engaging.

9. **Are instructions provided when you need them?**
   Instructions are provided for all aspects of the web site.
   Demonstrations are clear.

The graphs in figures 14-16 indicate the responses of three different cohorts: a) end-users, b) all the experts, and c) competent experts. The graphs capture not only the differences between the end-user and expert responses, but also the difference between the differences when the responses of all of the experts were compared to the responses of the experts who indicated their competence.

Comparison of the Expert and End-user Process Analysis

Figure 14 is a graphic representation of the comparison between the responses of the experts to the Simpson Educational Web Site Evaluation Tool and the responses of the end-users to the modified Simpson Educational Web Site Evaluation Tool. The responses were based on the content analysis of the web site www.bodymatters.com.

A t test determined that the responses of the end-users were significantly different from the responses of all the experts. The results of the t test indicated that t observed (3.29) was greater than t critical (2.45) at 0.05 degrees of freedom.

A t test determined that the responses of the end-users were significantly different from the responses of the competent experts. The results of the t test indicated that t observed (9.92) was greater than t critical (2.45) at 0.05 degrees of freedom.

A third t test determined that there was a significant difference between the responses of all of the experts and the competent experts for the content analysis of the web site www.bodymatters.com. The results of the t test indicated that t observed (6.62) was greater than t critical (2.45) at 0.05 degrees of freedom.

Summary. The results indicated that the end-users provided a significantly more positive evaluation than the expert adjudicators for the content analysis section. The end-users provided less positive responses to only three of the seven questions.

Figure 15 is a graphic representation of the comparison between the responses of the experts to the Simpson Educational Web Site Evaluation Tool and the responses of the

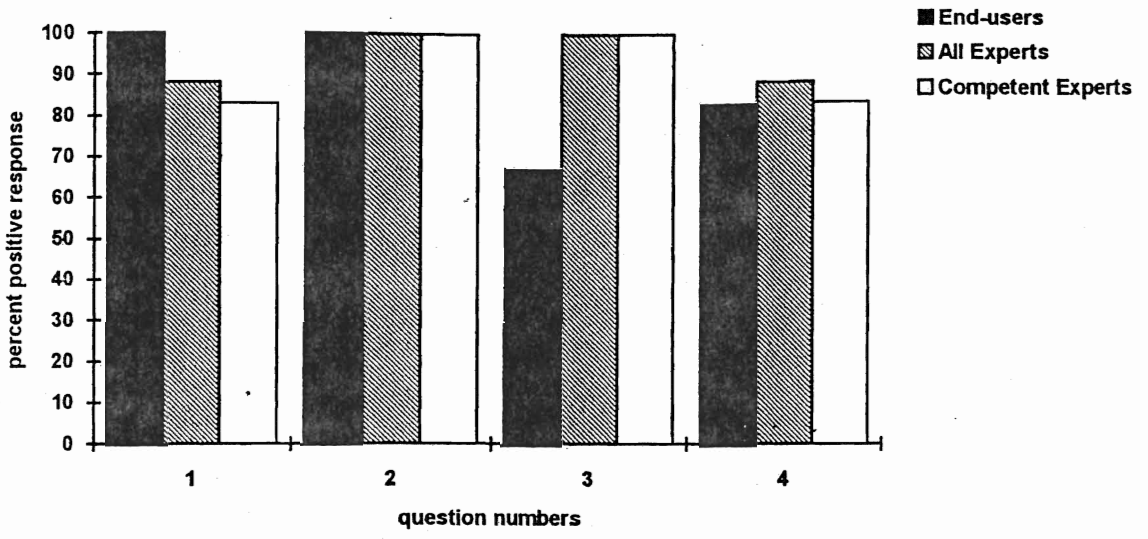


Figure 16. Comparison of the Expert and End-User Process Analysis

1. **Is it easy to move around the web site?**
   Menus are provided to make the web site user friendly.
   Each section is labelled with a heading.
   Image maps are provided.

2. **Do the pages of the web site load quickly?**
   The homepage loads smoothly and efficiently.
   There is efficient transition between screens.

3. **Do all of the pages download?**
   All of the pages download.

4. **Did you get lost in the web site?***
   Menus are provided to make the web site user friendly.
   Each section is labelled with a heading.
   Image maps are provided.

end-users to the modified Simpson Educational Web Site Evaluation Tool. The responses were based on the style analysis of the web site www.bodymatters.com.

A t test determined that the responses of the end-users were significantly more positive than the responses of all of the experts. The results of the t test indicated that t observed (2.95) was greater than t critical (2.45) at 0.05 degrees of freedom. It is important to note that the responses to questions 2 and 3 were not included in the calculation of t observed because the experts did not provide a response.

A t test determined that the responses of the end-users were significantly different from the responses of the competent experts. The result of the t test indicated that t observed (3.34) was greater than t critical (2.447).

A third t test determined that the responses of the competent experts were significantly different from all of the expert responses. The results of the t test indicated that t observed (2.54) was greater than t critical (2.45) at 0.05 degrees of freedom.

The results in Figure 15 show that the t test proved that the end-user provided a significantly more positive evaluation than the expert adjudicators for the style analysis section.

Figure 16 is a graphic representation of the comparison between the responses of the experts to the Simpson Educational Web Site Evaluation Tool and the responses of the end-users to the modified Simpson Educational Web Site Evaluation Tool. The responses were based on the process analysis of the web site www.bodymatters.com.

A t test determined that the responses of the end-users were not significantly different from the responses of all of the experts. The results of the t test indicated that t observed (0.68) was less than t critical (2.45) at 0.05 degrees of freedom.

A t test determined that the responses of the end-users were not significantly different from the responses of the competent experts. The results of the t-test indicated that t observed

(0.34) was less than t critical (2.45) at 0.05 degrees of freedom.

A third t test determined that the responses of the competent experts were not significantly different from the responses of all of the experts. The results of the t test indicated that t observed (0) was less than t critical (2.78) at 0.05 degrees of freedom.

The results indicated that the end-users and the experts provided a consistent and positive evaluation of the process analysis of the web site www.bodymatters.com.

Summary of the Categorical Analysis of the End-user Comments Section

Introduction

Three main categories emerged from a categorical analysis of the end-user comments section. The comments were qualitative data in which the end-users provided their opinion of the web site www.bodymatters.com. A comments section was provided to ensure that the end-user was not limited by the criteria in the End-user Educational Web Site Evaluation Tool. The end-users offered favourable opinions, unfavourable opinions, and suggestions for how to improve the web site. These three main categories were further subdivided into three sections; content, process, and style.

Content. The positive comments within the content category indicated that the end-users found the web site to be informative and interesting. The negative comments within the content category indicated that the information focused too much on the female changes during puberty.

Comprehension was an area that was problematic as indicated by the end-user comments. As one student indicated, "...alot of it is boring because I don't understand half of it." Other students indicated that spaces between words made comprehension difficult.

Process. There was one positive and one negative comment within the process section. A student responded that the web site was very easy to access, whereas another student voiced the opinion that the information for the males was difficult to access.

Style. The comments that fell within the style category were only positive. The students provided positive comments on the graphics, the layout, and the style with which information was delivered.

Summary of the Categorical Analysis of the Expert Comments Section

Introduction

Three main categories emerged from the categorical analysis of the expert comments section. These categories were favourable opinions, unfavourable opinions, and suggestions for how to improve the web site. These three main categories were further subdivided into three sections: content, process, and style.

Content. The Health Educator and the Teacher explored many areas within the content section and agreed that the "exercise" section was very good. The Health Educator did not identify any other areas that were good within the content category. The Teacher indicated that the web site was very informative. It is interesting to note that the Computer Scientist offered no comments for the content category.

The Health Educator and the Teacher agreed that there were parts of the web site that were too difficult for a grade 7 and 8 class. Both the Health Educator and the Teacher identified areas where the content was inaccurate or not comprehensive. The Health Educator identified that the corporate influence of Tampax played a role in the breadth and depth of the content. The Teacher recognized that the links were very commercialized and not educationally sound.

Style. The experts did not provide any positive comments about the style of the web site. The Teacher and the Health Educator identified that the web site was strictly informative and that it did not engage the students in the information. The Health Educator commented that the web site would lose the interest of adolescents, and the Teacher noted that the web site lacked visual support and interactivity. The Health Educator found that

the flow of the text was not very good, and that the tone in some of the sections was not appropriate to reassure the girls about the changes happening to their bodies.

The Computer Scientist noted that some of the pages had imposing backgrounds that made the information difficult to read. This expert also indicated that some of the pictures were just for aesthetic purposes and did not serve to enhance the educational value of the content presented.

Process. The experts offered no comments.

Summary of Expert Suggestions

Content. The educator suggested expanding the information on exercise routines and explaining some of the language in more detail. The Health Educator and the Computer Scientist had no suggestions for the content section.

Style. The Computer Scientist suggested that the background should be less obtrusive and that interactivity should be increased through on-line evaluations. The Health Educator and the Teacher had no suggestions for the style section.

Process. The only expert to speak to the process category was the Computer Scientist. This expert suggested e-mail links and posting sections at the bottom of each relevant page for the purpose of increasing interactivity. Another suggestion was to reduce the latency time by "name tagging" the separate sections. Finally, the Computer Scientist recommended the use of image maps to improve the delivery.

Interpretation of the Findings

The web site www.tampax.com was evaluated using two different approaches by two different cohorts. The unique results of the product-oriented approach and the performance-oriented approach will be described for their individual merit, and the relationship between a) the expert, product-oriented evaluation, and b) the end-user, product-oriented evaluation, and c) the end-user performance appraisal will be examined.

The End-user Performance Appraisal.

The performance appraisal, although limited by a small population, did demonstrate a trend of improved performance due to the intervention of the web site. As recommended by Mauldin (1996), the empirical evidence was gathered as a method of determining the instructional effectiveness of the courseware.

Performance appraisals were deemed important by Gill, Dick, Reiser and Zahner (1992) and Reiser and Dick (1990) when they determined that software that was rated highly effective by those using subjective evaluation techniques proved not to be highly effective when applied by the learners. Jolicoeur and Berger (1988) determined that teacher's subjective ratings of software programs were not valid indicators of the instructional effectiveness of the software.

The findings of the performance evaluation for the web site www.bodymatters.com established similar results. The Computer Scientist returned an overall score of 25, the Teacher returned an overall score of 13, and the Health Educator returned a score of -29 of a possible range of -132 to +132 for the Simpson Educational Web Site Evaluation Tool. These scores indicate that the experts reported a poor evaluation of the web site, yet the performance of the students improved. The finding that a subjective expert evaluation is not consistent with the effectiveness of the courseware is consistent with the findings of Jolicoeur and Berger (1988), Gill, Dick, Reiser and Zahner (1992) and Reiser and Dick (1990)

The reader should be cautioned of the significance of these results. The results of the performance appraisal indicate that performance improved; they did not indicate that learning took place. Future research should impose retention tests to determine if learning improved significantly. Another valid indicator of learning might be to compare the results of a class who learned through Internet-based courseware and a class that learned through traditional methods. The results of the present study must be considered

with the limitations of a pretest/posttest design. Testing was a threat to the validity of the findings because the completion of the pretest could have been the impetus to learning the correct answers to the posttest. A nonrandom population with little cultural or racial diversity also affects the importance of the results.

The Expert, Product-oriented Evaluation

Rowland (1994) noted that criterion-based checklists did not provide an overall assessment of the product and they did not evaluate the effect of the product on the learner. These problems emerged in the current research.

In the present study, as in the study by Rowland (1994), the sum of the responses did not provide a comprehensive assessment of the product. They did, however, identify the presence or absence of specific considerations (Sorge et al.,1993). These findings support the findings of the present study. The Simpson Educational Web Site Evaluation Tool did not provide an overall statement of the courseware, but its use did identify the presence or absence of specific criteria. Specifically, the experts returned a poor evaluation of the web site www.bodymatters.com for all sections except the social considerations section in the content analysis, and the teaching considerations and system efficiency sections of the process analysis, for which they returned a moderate evaluation. The separate sections of the Simpson Educational Web Site Evaluation Tool generated a report that examined specific aspects of the web site. Unfortunately, this report did not provide specific information as to why the experts provided a poor evaluation of the web site www.bodymatters.com.

Brill Pisik (1997) conducted a similar study on the product-oriented evaluation of training software. The author used a criterion-based checklist to produce a numeric score for the training software. Each statement was scored, the scores of each section were totaled, and the overall score was determined. Brill Pisik indicated that the numeric data allowed for a consistent comparison amongst the software packages because of the

numeric data and the use of the same evaluation tool. The findings of the current study contradicted the findings of Brill Pisik. Although the experts responded to the same criteria, their responses were not consistent for a number of reasons. Some of the experts responded to criteria they had no competence in, the experts did not examine the same parts of the web site as indicated by divergent responses, and different evaluators did not consistently interpret the criteria. These limitations will be discussed in the following paragraphs.

One of the significant findings that emerged from the research was that the weighting of the expert responses based on their competency was potentially more important than the weighting of the individual criterion. The results indicated that the experts frequently responded to criteria that they did not have competence in, as determined by self-reported data. A statistically significant difference was found between the responses of all of the experts and the responses of the experts who indicated competence in specific criteria. This was an important finding, since most criterion-based checklists do not solicit the expert's competence on individual criteria, but typically ask for expert competence in the general topic.

The accuracy of the responses was questionable as it was obvious that the adjudicators overlooked some aspects of the web site. For example, the expert adjudicators unanimously responded that the identity of the author was not provided. When the criterion was verified, it was determined that the identity of the author was indeed provided. The credibility section of the evaluation tool was plagued by inaccurate responses as the adjudicators provided inaccurate responses for four of eight criteria.

Another limitation to criterion-based checklists that emerged in the current research was that the interpretation of the criteria was not consistent between experts. To control for this limitation the author included a description of the criteria that required interpretation. The inclusion of the criteria description was an attempt to increase the

interrater consistency by ensuring that the criteria were defined. Unfortunately, as indicated by some of the results, the inclusion of the definition sheet did not ensure that all of the evaluators had a consistent understanding of the criteria. These results suggest the need for thorough training of the adjudicators for the purpose of increasing interrater reliability.

Northrup (1995) indicated that the validity of the evaluation tool was of utmost importance in ensuring a valid assessment. For the purpose of the current research the evaluation tool was rigorously critiqued by 20, fourth-year Health Studies majors who were taking a course entitled "Developing Electronic Curriculum." The validity of the tool was limited by the expertise of these evaluators. One problem with the Simpson Educational Web Site Evaluation Tool that emerged from the research was that the author did not consider the option of an expert indicating "Not Applicable."

The problems associated with Likert scale data should also be examined. The degree to which the experts agreed with each criterion is in itself a subjective indicator, which removes any attempt to show an objective examination. Furthermore, due to the fact that the experts were asked to indicate their competence, it was difficult to run a statistical test on an incomplete data set. This problem was confounded by the fact that some of the experts responded to criteria that they indicated they did not have competence in.

Some of the literature on criterion-based checklists indicated that it is difficult to know how to weight the importance of individual criteria. This was also a problem in the current research where each of the criteria were weighted equally. This is a limitation to the current study where such criteria as "The information is accurate" is weighted equally with "The author provides contact information beyond an e-mail address." These criteria demonstrate the distinction between a criterion that is necessary and a criterion that

would augment the site if it were there, but would not necessailry significantly effect the web site.

There was also a difference in the type of questions. The checklist contained questions that were objective and easily verified such as "The identity of the author is provided" for which there is a correct, concrete response, versus "The content is engaging" which is a subjective measure. Perhaps future research could examine the relationship between accurate responses to the objective criteria, and how that compares with the subjective criteria.

Other factors relating to the experts themselves can also change the results of a criterion-based checklist. Motivating factors such as boredom, mood, the desire to please by providing positive responses, personal style of criticism, and responses to the corporate aspect of the web site could have affected the responses of the experts.

Finally, other problems with expert evaluation were identified by Aedo et al. (1996) and Reiser and Kegelmann (1994). These authors indicated that criterion-based checklists produced biased evaluations due to the orientation of the expert, and inauthentic evaluations because the end-user did not have the opportunity to express their point of view. These limitations were overcome in this project, as experts from three different orientations were used to solicit diverse orientations, and the end-users were given the opportunity to evaluate the web site www.bodymatters.com from their orientation.

The End-user Application of a Criterion-based Checklist

The end-user criterion-based evaluation was not plagued by as many problems as the expert criterion-based evaluation. The majority of the end-users provided a positive and consistent assessment. The end-users responded to a simplistic checklist that was adapted from the expert criterion-based checklist, and the qualitative responses generated from Pilot Study 1.

The findings of the current research were consistent with the findings of Reiser and Kegelmann (1994), who found that teachers and students often rated software quite differently. The current courseware evaluation produced results in which the end-users provided a significantly more positive evaluation than the expert adjudicators, as determined by a t test.

A significant finding in the current research was that the students did not provide qualitative comments regarding the processes of the web site. This appears to be an area to which the end-user adjudicators' attention must be drawn. The criterion-based checklist elicited the students' opinion of the web site processes.

There appeared to be some concern about the qualifications of end-users to critique courseware. Jolicoeur and Berger (1988) did not recommend the use of students as evaluators as they were not able to accurately judge the instructional quality of the courseware. This could be because the end-users were not accustomed to critically evaluating courseware that was designed for them (Reiser and Kegelmann, 1994).

Conversely, Smith and Keep (1986) found that children were sophisticated in their judgment of the courseware. The current end-user evaluation indicated that the end-users provided a consistent evaluation based on a simplistic criterion-based checklist.

Comparison between the Expert and the End-user Product-oriented Evaluation

The findings of the present study were comparable to the findings of Reiser and Kegelmann (1994), who found that when subjective evaluation methods were used, the students and teachers rated the curriculum differently. The results of the comparison between the checklist data and the comments data revealed significant differences between the experts and the end-users.

The results indicate that the end-users provided a significantly more positive evaluation than the expert adjudicators for both the content and style analysis, but that there was no significant difference between the groups for the process analysis. The

process analysis was rated positively by both the experts and end-users; however, there was limited consistency between the comparison of the content and style analysis of the web site.

The comparison of the results of the checklists confirmed many of the findings of the two cohorts. The end-users and the experts indicated that the reading level was inappropriate to the target audience, that the web site was well laid out, that it was easy to move around the web site, and that the pages all download quickly. The process analysis was very consistent across the two groups. This is likely because the experts and the end-users did not require any special competencies to recognize whether or not they got lost in the web site or if the pages download.

These findings were limited by the validity of the compared criteria. It is possible that the compared criteria are not sensitive enough to capture the same concept. The experts used multiple criteria to capture the same concept that the end-user responded to.

The findings were also limited by the effect of the intervention on the end-users. It was possible that the end-users thought that they had to return a positive evaluation of the web site. Students are not often given the opportunity to critique courseware, and it was possible that some of the students were either not comfortable with the experience, or they lacked the critical thinking skills required to evaluate the courseware. It is also possible that the positive responses were an indicator of the end-users being excited to be away from their classroom, in a university setting, with the opportunity to work on the Internet. Conversely, it is also possible that the students provided a valid evaluation from their perspective based on a simple criterion-based checklist that was age and language appropriate.

Expert Comments Section

The literature indicated that the primary limitation of soliciting qualitative comments data was that it was time consuming (Aedo et al., 1996). For the purposes of

the current study the qualitative data provided more in-depth information than what was garnered from the checklist data. It appeared that a union between the two methods was important to ascertain a valid evaluation. The criterion-based checklists brought the attention of the experts to aspects of the web site that they might not have considered on their own. The comments section allowed adjudicators the freedom to embellish or provide insight into relevant points that the checklist might have overlooked.

An interesting finding that emerged from the comments section was that the experts tended to comment only on issues within their own area of expertise. This was an important consideration when choosing the experts to evaluate Internet-based courseware. It was important to choose experts who provided an in-depth examination of pertinent features.

The experts also offered a number of suggestions for improving the web site, which was information not garnered by the criterion-based checklist. It was also interesting to note that some of the experts approached the evaluation only as a critique and did not offer any positive feedback or suggestions (i.e. health educator)

End-user Comments Section

The comments provided by the end-users supported the results of the end-user Educational Web Site Evaluation Tool. The end-users provided a positive evaluation of the content, style and process of the web site, and their comments reaffirmed this and served to specify the areas that the end-users were identifying in the checklist.

The comments of the end-users were not as comprehensive as those of the experts, but they did serve to identify two key weaknesses of the web site that the expert evaluators failed to identify. The findings of the current research were consistent with the findings of Sorge et al. (1993). The author found that end-user input was a valuable part of the evaluation process because the end-users identified things that the experts overlooked. In the present research, the end-users identified that there was not enough

information directed at boys and puberty, and the experts did not. The end-users also identified that the spacing between some of the words limited their comprehension, yet the experts overlooked this.

An interesting finding was that the end-users provided only two comments to the processes of the web site, and they did not speak to the processes of the web site in Pilot Study 1. This was a strong indicator that the students' attention must be drawn towards this area to complete a comprehensive assessment.

Comparison Between the End-user Comments and the Expert Comments

The differences between the results of the experts and the results of the End-users outweigh the similarities. The similar findings between the two cohorts were that the web site was very informative, but that the language used was too difficult.

The differences between the end-users and the experts were numerous. The end-users liked the style of presentation, while the experts did not. The majority of the end-users' comments focused on the style of the web site, which was a finding similar to that of Smith and Keep (1992). The end-users commented on the processes of the web site where the experts did not.

The students were unable to speak to the comprehensiveness of the web site, nor were they able to judge whether or not they were being exposed to misinformation. This was because they did not have any expertise in the content. The content analysis section was the main section where the end-users were unable to provide a valid evaluation.

There are important exceptions to this however. It should be noted that the experts did not identify that the information was mainly directed to problems and issues that young women would be encountering. The end-users identified that the split-up words made comprehension difficult for them. Age and concept appropriateness were something that experts could only assume.

CHAPTER FIVE: SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Summary of the Research

Although the number of educational web sites posted to the Internet continues to grow rapidly, no standardized evaluative process is used to filter this information. This is problematic for the application of the Internet as a tool in education. There are many ways to evaluate Internet-based courseware; however, current related literature is inconclusive in establishing guidelines for the evaluation of Internet-based courseware. The purpose of the current research was to assess evaluative measures for Internet-based courseware. Specifically, two entities were evaluated within the study: a) the outcome of the product, and b) the product itself. The main objective of the research was to contribute to the knowledge that will be required to develop standardized Internet-based courseware evaluation guidelines for educators.

To this end, the web site www.bodymatters.com was evaluated using two different approaches by two different cohorts. The first approach was a performance appraisal by a group of end-users. A positive, statistically significant change in the students' performance was observed due to the intervention of the web site www.bodymatters.com.

The second approach was a product-oriented evaluation of the web site www.bodymatters.com with the use of a criterion-based checklist and an open-ended comments section. The first cohort to complete the product-oriented evaluation was a group of experts who consisted of a computer scientist, a health educator and a classroom teacher. These experts applied the Simpson Educational Web Site Evaluation Tool to the web site www.bodymatters.com. The results of the criterion-based evaluation indicate that the experts returned a poor evaluation of the web site www.bodymatters.com. A significant finding was that there was limited consistency among the expert responses. The limitations of the criterion-based checklist were revealed, and it was found that the

self-reported competence of the experts played a large role in determining the validity of the report.

The expert comments indicated that they commented only on aspects of the web site that fell within their area of expertise, even though their attention was drawn to the other salient features of the web site by the Simpson Educational Web Site Evaluation Tool. The comments provided a better picture of the web site than did the criterion-based checklist.

The second cohort to conduct the product-oriented evaluation was a group of grade 7 students from a Niagara region public school. These end-users applied the End-user Educational Web Site Evaluation Tool to the web site www.bodymatters.com. The results of the criterion-based evaluation indicated that the end-users provided a positive and consistent evaluation of the web site www.bodymatters.com.

The end-user comments indicted that the end-users were unable to speak to the processes aspects of the web site, and that the end-users commented on aspects of the web site that the experts overlooked.

The comparison between the expert and end-user product-oriented evaluation results showed that the end-users provided a significantly more positive report for the content and the style of the web site, and that both cohorts returned a positive evaluation of the process analysis. The experts were best equipped to critique the content of the web site, but the end-users commented on some things that the experts did not.

Overall, the performance appraisal by the end-users indicated that the web site was effective in improving the students' performance. The expert, product-oriented evaluation returned a poor rating of the web site, and the end-user, product-oriented evaluation returned a good rating of the web site.

Conclusions

The findings of the current research will contribute to the development of a standardized method for evaluating Internet-based courseware. The findings will also provide guidance for educators who are interested in introducing the Internet into their classrooms.

The problem is twofold: a) The many uses of the Internet in education are well established in the related literature, yet there is no established standardized method for evaluating Internet-based courseware, and b) educators do not participate in the development of Internet-based courseware, yet they are encouraged to use it in their classrooms. These problems create a need for standardized, summative evaluation methods for Internet-based courseware that can be implemented by the educator.

The findings indicate that there is a significant difference between the product-oriented checklist evaluation submitted by the experts and the end-users. There was also limited consistency amongst the responses of the experts from different orientations. The experts commented only on issues within their own area of expertise, and there was a significant difference between the results of the experts who rated themselves competent on certain criteria versus experts who did not rate themselves as competent. The results also indicated that the experts did not consider themselves competent to answer all of the criteria, which raises the concern that one person is not competent to return a complete checklist. Based on the comparison between the questions that the three different experts felt they could not speak to, it would appear that a summative, criterion-based evaluation can be completed only by a multidisciplinary team. This finding complements the formative evaluation research, which also calls for a multidisciplinary team.

The end-users commented on issues that the experts overlooked, but were unable to provide sound evaluation of the actual content of the web site. The end-users returned a positive evaluation of the web site.

As indicated in previous research, expert subjective evaluations that rated courseware poorly did not indicate that the efficiency of the product would be limited. This supports Rowland's (1994) research that found that the sum of the different criteria did not add up to the sum of the web site. The current research confirmed this in that the experts returned a poor report of the web site, yet the end-users' performance improved due to the intervention of the web site.

Implications for Practice

The purpose of the current research is to provide educators with guidelines for the assessment of Internet-based courseware. The in-depth evaluation conducted for the purpose of this research would be too time consuming for an educator, therefore the following guidelines were modified from the optimal evaluation techniques to provide a practical approach for an individual educator.

Guidelines for Educators

1. Determine the objective of your evaluation. (Is it to meet the needs of the curriculum, to provide a good learning experience for the students, to teach the students about a particular topic?) The objective will guide the evaluation.

2. Use a criterion-based checklist for the purpose of identifying the presence or absence of specific criteria. This will ensure that all salient features of the web site are considered. Ensure that the checklist is subdivided into subsections to facilitate this process.

3. Maintain the checklist information for future reference and comparison.

4. Write a summary of the strengths and weaknesses of the web site.

5. Based on the results of the criterion-based checklist, determine if the courseware meets the objectives determined in stage one. If the courseware has passed the initial inspection it can be incorporated as a trial in the classroom.

6. Plan an educational experience for the students by inviting them to critically

review the web site using the End-user Educational Web Site Evaluation Tool. This will allow the students to examine the web site and to practice critical thinking skills.

7. Invite the students to comment on the web site.

8. Compile the scores of the students on the End-user Educational Web Site Evaluation Tool. Compare these scores to the educator's score and determine areas of divergence.

9. Read over the students' comments section to determine if their evaluation reveals anything that the educator overlooked.

10. Determine, based on the evidence thus far, if the web site has met with a positive evaluation. If it has met with a positive evaluation, incorporate the web site into the curriculum, and share the good web site with fellow educators.

If the educator is interested in taking the evaluation one step further:

11. If possible, teach one class in the traditional method, and teach the other class based on the Internet-based courseware.

12. Measure the performance success by comparing the grades to the traditional classroom, or create a pretest and a posttest.

Implications for Theory

The findings of the current research will contribute to the development of a standardized method for evaluating Internet-based courseware The rigorous evaluation of the current web site has contributed some important findings for a summative review of educational web sites.

1. The current research found that although the web site received a poor subjective rating by experts, the performance appraisal determined that it was an effective piece of courseware. This finding was comparable to the research on

computer software assessments conducted by Reiser and Dick (1990) and Gill, Dick, Reiser and Zahner (1992).

2. The findings of the current research are similar to the findings of Rowland (1994) who found that criterion-based checklists do not provide a good overall picture of the courseware.

3. The current research confirmed the findings of Sorge et al. (1993) who found that end-users provide insight into aspects of a web site that experts overlook.

4. The current research found that experts from different orientations comment on findings within their area of expertise. These findings indicate that it is important to choose experts from different orientations who can meet the goal of the evaluation.

5. Experts and End-users provide limited comments to the processes of the web site. The findings indicate that the adjudicator must be made aware of this consideration with the use of a questionnaire or checklist.

6. End-users and experts returned significantly different criterion-based evaluations for the content and style aspects of the web site, but returned similar evaluations for the process aspect of the web site.

7. Experts who rated themselves competent in a specific criterion returned a significantly different evaluation than all of the experts who responded to the criterion.

Implications for Further Research

The implications for this research in practice are numerous. If an effective, standardized methodology emerges from this research, the educator will be able to apply this evaluation procedure within their learning environments. The findings will

contribute to the knowledge required to develop guidelines for educators who will be using Internet-based courseware in their learning environments.

Future research should consider the following:

1. Conduct a retention test for the performance appraisals. This will determine if learning took place instead of just an improvement in performance.

2. Compare the effectiveness of the courseware with a traditional method of teaching to determine which method was better at facilitating learning.

3. Conduct comprehensive training of adjudicators who will be completing the criterion-based checklist. Have the evaluators evaluate one piece of courseware and discuss the results. Check the consistency of expert ratings on simple web sites to ensure proper completion of the checklist.

4. Use a multi-disciplinary team of experts to conduct the evaluation. Do not expect to have consistent results for the evaluation because each expert will focus on own areas of expertise.

5. Determine the goal of the evaluation.

6. Weight criteria according to importance.

7. Ensure that specific questions about processes of the web site are provided to ensure that the processes are evaluated

8. Use a criterion-based checklist that is divided into sections to help determine general areas of strength or weakness.

9. Invite comments from your adjudicators. This can provide an evaluation that is more in-depth.

References

Aedo, I., Catenazzi, N., & Diaz, P. (1996). The evaluation of a hypermedia learning environment: The CESAR experience. Journal of Educational Multimedia and Hypermedia, 5(1), 49-72.

Alexander, J. & Tate, M. (1996) Checklist for an informational web page. **http://www.science.widener.ed./~withers/inform.htm** Wolfgang Memorial Library. Widener University. [24 October 1996].

Brill Pisik, G. (1997, July-August). Is this course instructionally sound: A guide to evaluating online training courses. Educational Technology, 50-51.

Collis, B. (1996, November-December). The Internet as an educational innovation: Lessons from experience with computer implementation. Educational Technology,. 21-30.

Cook, T. D., & Campbell, D. T. (1979). Quasi-experimentation: Design and analysis issues for field settings: Boston: Houghton Mifflin.

Crosby, M. E., & Stelovsky, J. (1995). From multimedia instruction to multimedia evaluation. Journal of Educational Multimedia and Hypermedia, 4(2/3), 147-162.

Doll, C. A. (1987). Evaluating educational software, Chicago IL: American Library Association.

Flagg, B. N. (1994). Learning science from children's radio: Summative evaluation of Kinetic City Super Crew. Educational Technology Research and Development, 42(3), 29-43.

Flake, J. L. (1996). The World Wide Web in education. Computers in the Schools, 12(1/2), 89-100.

Gill, B., Dick, W., Reiser, R.A., & Zahner, J.E. (1992). A new model for evaluating instructional software. Educational Technology, 32(3), 39-44.

Gros, B., & Spector, J. M. (1994, May-June). Evaluating automated instructional design systems: A complex problem. Educational Technology, 37-46.

Hackbarth, S. (1997, May-June). Integrating web-based learning activities into school curriculums. Educational Technology, 59-65.

Hannafin, M. J., Hall, C., Land, S., & Hill, J. (1994, October). Learning in open-ended environments: Assumptions, methods, and implications. Educational Technology, 48-54.

Harper, S. (1988, fall). A model for program evaluation. Education Canada, 18-23.

Harris, R. (1997) Evaluating Internet research sources. **http://www.sccu.edu/faculty/R_Harris/evalu8it.htm** [07 July 1997].

Hoadley, M. R. (1996). Use of technology in health instruction. Journal of Health Education, 27(5), s44-s47.

Holzschlag, M. E. (1997). Sizzling web site design (L. Lemay, Ed.). Indianapolis, IN: Sams.net .

Jolicoeur, K., & Berger, D. E. (1988). Implementing educational software and evaluating its academic effectiveness: Part I. Educational Technology, 28(9), 13-19.

Kearsley, G., Lynch, W., & Wizer, D. (1995, November-December). The effectiveness and impact of online learning in graduate education. Educational Technology, 37-42.

Kirk, E. E. (1996)Evaluating information found on the Internet. **http://milton.mse.jhu.edu:8001/research/education/net.html** [05 July 1996]

Kotecki, J. E., & Siegel, D. E. (1998). Use of a critical thinking/questioning approach to evaluate WWW information. American Journal of Health Behaviour, 22(1), 75-76.

Landow, G. (1992). Hypertext: The convergence of contemporary critical theory and technology. Baltimore: The John Hopkins Press.

Lasarenko, J. (1997). Wired for learning: Harness the power of the Internet for education. Indianapolis, IN: Que Corporation.

Lohr, L., Ross, S. M. & Morrison, G. R. (1995). Using a hypertext environment for teaching process writing: An evaluation study of three student groups. Education Technology Research and Development, 43(2), 33-51.

Maddux, C. D. (1994, September). The Internet: Educational prospects and problems. Educational Technology, 37-42.

Maddux, C. D., & Johnson, D. L. (1997, September-October). The World Wide Web: History, cultural context, and a manual for developers of educational information-based web sites. Educational Technology, 5-12.

Mauldin, M. (1996, March-April). The formative evaluation of computer-based multimedia programs. Educational Technology, 36-40.

McLauchlan, K. (1996) WWW cyberguide ratings for content evaluation **http://www.cyberbee.com/guide1.htm** [05 August 1996]

Monahan, B. D., & Dharm, M. (1995, January-February). The Internet for educators: A user's guide. Educational Technology, 44-48.

Muller, E. W. (1985). Application of experimental and quasi-experimental research designs to educational software evaluation. Educational Technology, 25(10), 27-31.

Neuman, D. & Marchionini, G. (1995). Evaluating Perseus 1.0: Methods and final results. Journal of Educational Multimedia and Hypermedia, 4(4), 365-382.

Northrup, P. T. (1995, November-December). Concurrent formative evaluation: Guidelines and implications for multimedia designers. Educational Technology, 24-31.

Ohl, T. M., & Cates, W. M. (1997, November-December). Applying metaphorical interface design principles to the World Wide Web. Educational Technology, 25-38.

Owston, R. D., & Wideman, H. H. (1987). The value of supplemental panel software reviews with field observations. Canadian Journal of Educational Communication, 16(4), 295-308.

Persico, D. (1997). Methodological constants in courseware design. British Journal of Educational Technology. 28(2), 11-123.

Rada, R., Rimpau, J., Bowman, C., Gordon, J., Henderson, T., & Sansom, T. (1996, September-October). World Wide Web activity and the university. Educational Technology, 49-51.

Rakes, G. C. (1996, September-October). Using the Internet as a tool in resource-based learning environments. Educational Technology, 52-56.

Rankin, W. (1997 July-August). The cyberjournal: Developing writing, researching, and editing skills through e-mail and the World Wide Web. Educational Technology, 23-31

Reiser, R. A., & Dick, W. (1990). Evaluating instructional software. Educational Technology Research and Development, 38(3), 43-50.

Reiser, R. A., & Kegelmann, H. W. (1994). Evaluating instructional software: A review and critique of current methods. Educational Technology Research and Development, 42(3), 63-69.

Rowland, G. (1994, January). Designing and evaluating: Creating futures and appreciating error. Educational Technology,10-22.

Schotsberger, P. G. (1996, March-April). Instructional uses of the World Wide Web: Exemplars and precautions. Educational Technology, 47-50.

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, et al.(Eds.) Perspectives of curriculum evaluation, (pp. 39-83). Chicago: Rand McNally.

Smith, D., & Keep, R. (1986). Children's opinions of educational software. Educational Research, 28(2), 83-88.

Sorge, D. H., Campbell, J. P., & Russell, J. D. (1993, April/May). Evaluating interactive video: Software and hardware. Tech Trends, 19-26.

Spotts, T. H., & Bowman, M. A. (1995, March-April, p.). Faculty use of instructional technologies in higher education. Educational Technology, 56-64.

Squires, D., & McDougall A. (1994). Choosing and using electronic software: A teacher's guide. Washington DC: Falmer Press.

Staninger, S. W. (1994, July-August). Hypertext technology: Educational consequences. Educational Technology, 51-53.

Stufflebeam, D. L., & Shinkfield, A. J. (1985). Systematic evaluation, Boston, Kluwer-Nijhoff.

Starr, R. M., & Milheim, W. D. (1996, September-October). Educational uses of the Internet: An exploratory survey. Educational Technology, 19-28.

Summers, J., Using the Internet to enhance teaching and learning, (1996). In C. C. Kuhlthau, M. E. Goodin, & M. J. McNally, The virtual school library: gateway to the information superhighway, (pp. 21-27), Englewood, CA: Libraries Unlimited.

Symons, A. K. (1997, April). Sizing up sites. School Library Journal, 22-25.

Weston, C., McAlpine, L. & Bordonaro, T. (1995). A model for understanding formative evaluation in instructional design. Educational Technology Research and Development, 43(3), 29-48.

Wilkinson, G. L., Bennet, L. T., & Oliver, K. M. (1997, May-June). Evaluation criteria and indicators of quality for Internet resources. Educational Technology, 52-58.

Willing, K. R., & Girard, S. (1990). Learning together: Computer-integrated classrooms. Pembroke.

Woolsey, K., & Bellamy, R. (1997). Science education and technology: Opportunities to enhance student learning. The Elementary School Journal, 97(4), 385-399.

Appendix A

The Original Simpson Web Site Evaluation Tool

DEFINING THE WEBSITE

- ☐ Is the topic defined clearly?*
- ☐ Is the purpose of the website stated?*
- ☐ Are the objectives of the website stated?*
- ☐ Is adequate & relevant information provided in order to meet the objectives?*
- ☐ Is the target audience defined?*
- ☐ Is the information accurate?*
- ☐ Is the information current?
- ☐ Was the last update recent?*
- ☐ Are spelling, grammar and punctuation correct?*
- ☐ Are the links relevant & appropriate to the topic?*
- ☐ Are the criteria for selecting the links explicit?*
- ☐ Does the content stand alone, or is it merely a list of links to other sites?

CREDIBILITY

- ☐ Is the identity of the author provided?
- ☐ Is information about the author's training and education provided?
- ☐ Is this training and education from a well-respected organization?
- ☐ Does the author provide contact information beyond just an e-mail address?
- ☐ Are references and a bibliography provided?
- ☐ Is the site subject to a review process?
- ☐ Does the company that is producing the website provide unbiased information?
- ☐ Are all aspects of the topic presented?
- ☐ Are there no omissions due to the agenda of the group who produced the website?

SOCIAL CONSIDERATION

- ☐ Are people of different races, ages, and physical and mental abilities pictured and mentioned?
- ☐ Are different groups presented in positive and non-stereotypical ways?
- ☐ Are issues specific to these groups in evidence?*
- ☐ Does the website avoid an obvious cultural bias?*
- ☐ Are the social messages in the content positive?*
- ☐ Does the website promote inclusion and acceptance?*
- ☐ Does the website offer positive role models?*
- ☐ Does the website help the student to develop decision-making skills?

TEACHING CONSIDERATIONS

- ☐ Is the focus of the website educational?*
- ☐ Does the website support the school curriculum?
- ☐ Is the reading level appropriate for the target audience?*

- ❑ Are the concepts age-appropriate for the target audience?*
- ❑ Are definitions provided when necessary?
- ❑ Is the content engaging?*
- ❑ Is the content comprehensive?

--Style Analysis—

DESIGN

- ❑ Are graphics and animation clear and easily interpreted?
- ❑ Are graphics and audio appropriate to the population?
- ❑ Does the visual display occupy most of the screen?
- ❑ Are the characters of a legible size?
- ❑ Is the sound used with purpose?*
- ❑ Is the format varied and interesting?

FEATURES

- ❑ Does the website have interactive features such as CGI scripts & conferencing?
- ❑ If calculations are required, can they be done onscreen?
- ❑ Are bookmarks and notepads provided for the students?
- ❑ Is clipping information and automatic citations possible?
- ❑ Are students denied access to confidential or inappropriate information?
- ❑ Are instructions provided for all aspects of the website?
- ❑ Does the student control the rate of everything presented?

NAVIGATION

- ❑ Are menus provided to make the website user friendly?
- ❑ Is each section labelled with a heading?
- ❑ Are image maps provided?
- ❑ Is a search engine available to the user?*
- ❑ Does the search engine index the entire website?*

TEACHING CONSIDERATIONS

- ❑ Is the pedagogic approach superior using this medium?*
- ❑ Is the pedagogy innovative?*
- ❑ Are instructional strategies appropriate?*
- ❑ Are demonstrations clear?
- ❑ Is minimal teacher supervision required?
- ❑ Is there opportunity to answer open ended questions?
- ❑ Can students hypothesize and experiment?
- ❑ Does the program provide a forum for student interaction?
- ❑ Are the time limits of the class addressed by providing workable units?
- ❑ Is the material organized and well sequenced?

TESTING

- ❑ Does the website provide criteria to assess learning?
- ❑ Do the questions in the test measure the mastery of the content?
- ❑ Do student responses match the program objectives?
- ❑ Is feedback provided that is positive, informative and timely?

Process Analysis

INPUT CONSIDERATIONS

- ❑ Does the website allow users to correct input errors?
- ❑ Does the website accept abbreviations?
- ❑ Are minor variations in spelling and answers allowed?
- ❑ Are control keys used consistently?
- ❑ Are both mouse & keyboard control allowed?

SYSTEM EFFICIENCY

- ❑ Does the homepage load contiguously?*
- ❑ Can the startup screen be bypassed?
- ❑ Is there efficient transition between screens?*
- ❑ Can the website run without any special software requirements?
- ❑ Can the text stand-alone if the graphics are not loaded?
- ❑ Are out of date links quickly removed?
- ❑ Do all of the pages download?

TEACHING CONSIDERATIONS

- ❑ Can the teacher change or add content?
- ❑ Can the website store the test scores of a number of students?
- ❑ Are the results easily accessible?
- ❑ Are teachers provided with user guides about the website?

SUPPORT

- ❑ Is a help function available to users at all times?
- ❑ Are the materials required to run the site defined?
- ❑ Are any support materials required, and if so, are they provided?

Appendix B

Questionnaire examining the Simpson Web Site Evaluation Tool

Questionnaire

1. Are there any questions which are not clear? Y N

2. If yes, please list the questions.

   _____

   _____

   _____

3. Did you think that all of the questions were relevant? Y N

4. If no, please list the questions that were not relevant.

   _____

   _____

   _____

5. Would you add another category of questions to the Content section? Currently the categories are Defining the Web site, Credibility, Social Considerations and Teaching Considerations.
   Y N_____

   _____

   _____

6. Would you add another category of questions to the Style section? Currently the categories are Design, Features, Navigation and Teaching Considerations and Evaluation.
   Y N_____

   _____

   _____

7. Would you add another category of questions to the Process section? Currently the categories are Input Considerations, System Efficiency, Teaching Considerations and Support
   Y N_____

   _____

   _____

8. Would you add a specific question?

9. Did you find this evaluation tool easy to use? Y N

10. This evaluation tool was designed for educators to evaluate internet curriculum. Do you think that this tool provides a valid evaluation of a web site? Y N

11. What was good about the web site evaluation tool?

12. Do you have any suggestions to improve this web site evaluation tool?

Appendix C

Results of the Questionnaire examining the Evaluation Tool

Results of the Questionnaire Examining the Original Simpson Educational Web Site Evaluation Tool

1. **Are there any questions that are not clear?**
   20 YES / 3 NO

Please list the questions that are unclear

Are image maps provided? (7)*
Are there omissions due to the agenda of the group that produced the web site? (6)
Is the pedagogy innovative? (6)
Is the pedagogy superior using this medium? (5)
Are the criteria for selecting the links explicit? (4)
Is the site subject to a review process? (3)
Does the web site run contiguously? (2)
Are bookmarks and notepads provided for the students? (2)
Are the materials required to run the site defined? (2)
Are any supporting materials required, and if so, are they provided? (2)
CGI scripts & automatic citations & conferencing? (2)
Are the time limits of the class addressed by providing workable units?
Does the student control the rate of presentation?
Is the content comprehensive?
Is a search engine available to the user?
Does the search engine index the entire web site?
Is the content engaging?
Are issues specific to these groups in evidence?
Are the materials organized and well sequenced?
Does the web site accept abbreviations?
Can the teacher change or add content?

2. **Did you think that all of the questions are relevant?**
   12 YES / 11 NO

Please list the questions that are not relevant:

Are the time limits of the class addressed by providing workable units? (3)
Does the web site offer positive role models? (3)
Do the student responses match the program objectives? (2)
Does the web site help the students to develop decision-making skills?
Is minimal teacher supervision required?
Is the sound used with purpose?
Are bookmarks provided?
Are there no omissions due to the agenda of the group who produced the web site?
Are the social messages in the content positive?
Is there efficient transition between screens?
The entire Social Considerations section

* The number in brackets indicates the number of students with the same response.

3. **Would you add another category of questions to the *Content* section? Currently the categories are Defining the Web Site, Credibility, Social Considerations and Teaching Considerations.**
   18 NO / 5 YES

Please list the category of questions you would add:

Disclosure for the purpose of providing information on owners of site, and sponsors, advertisements and financial support of site should be visible and prominently displayed on the page.
Quality
Accessibility
Evaluator Considerations
1) What is your profession?
2) Are you the person in your organization that does the most evaluation of web sites?
3) Within a given period how often do you use the tool?
Layout and Organization of the content or page
Quantity

4. **Would you add another category of questions to the *Style* section? Currently the categories are Design, Features, Navigation and Teaching Considerations and Evaluation.**
   18 NO / 5 YES

Please list the category of questions you would add:

Graphics
(1) Useful?
(2) Relevant?
Quality of style
Links Evaluation
(1) Do they augment the information in the site?
(2) Are links used to fill in missing areas of a site?
(3) Do links take users to areas complementary to the site?
(4) Are the links easy to access?
(5) Do the links keep the user on the homepage?
Creativity & Innovation
Language

5. **Would you add another category of questions to the *Process* section? Currently the categories are Input Considerations, System Efficiency, Teaching Considerations and Support**
   21 NO / 2 YES

Learning Considerations
Quality of Process

6. **Would you add a specific question?**
   13 YES / 10 NO

Please list the question you would add:

Does the Web Site enlighten you to new and up to date information?
Is the content repetitive?
Is a disclaimer present?
Is there incorporation of new content?
Does content present itself as unique and consistent?
Does the site maintain the use of new and innovative technologies?
Do graphics pertain to the topic?
Are the interactive features effective in engaging the user?
Is the use of color used effectively?
Is the layout aesthetically pleasing?
How much experience does the evaluator have with computers?
How is the quality or usefulness of videos?
Does the Web Site offer something new that cannot be found in other Web Sites?
Is the Web Site beneficial to the topic?
Does the Web Site make people want to come back and use the Web Site as a reference?
Does the Web Site make people want to continue using the Internet?
Can this page be located without having the address?
Is the page linked to other Web Sites?
Is the speed of loading quick enough?
Is the response time quick to a request for support?
Are buttons available to make navigation through the site a smooth continuous journey, or does the user have to continually return to a menu to advance to the next screen

9. **Did you find this evaluation tool easy to use?**
   17 YES / 6 NO

If the evaluation tool was difficult to use, please list why:

too long & tedious (5)
some questions are not clear (5)
questions weren't relevant for every site (2)
questions can be looked at in many ways
questions too wordy
questions may be beyond the scope of the evaluator
questions that required computer knowledge
need space for comments (2)
unsure if boxes were for yes or no
put it on a computer or disk

If the evaluation tool was easy to use, please list why:

very thorough (2)
well laid out & broken into categories (2)
in depth & detailed
good range

10. This evaluation tool was designed for educators to evaluate Internet curriculum. Do you think that this tool provides a valid evaluation of a Web Site?
20 YES / 1 NO / 2 UNDECIDED

Comments

I think it was a little difficult to use because most sites are used to provide information, most don't test or record results. If there was some common educational site browser then we could evaluate these sites.

The experience of the evaluators and the frequency with which the evaluator uses the tool will have a direct effect on the results. Is training necessary to ensure consistency?

I think it begins a critical thinking process and that upon evaluation of the evaluation tool a valid evaluation can be made.

I think there was more focus on evaluating the appearance of this site rather than the evaluation of the content and its relevance and value to the educators and the public.

Undecided because it is from the educators perspective. There should be more questions related to the learner's perspective. Is it easy to understand? Is it fun and informative? Are there interactive areas such as games, charts etc. to assist in learning?

The evaluation covers issues important to educators.

The tool touches on all of the necessary elements needed for an effective, easy to use curriculum Web Site.

It is very comprehensive and covers everything.

It is a good tool to evaluate Internet courseware because it looks at a wide range of aspects.

This tool is valid because it touches on the main points of the Web Site. Content, style and process. Each area is subject to an intense critical analysis with many in depth questions.

11. What was are some of the strengths of this Web Site evaluation tool?

It is thorough & comprehensive (10)
It is well organized into subgroups & categories (5)
Precision, volume, depth, range of questions (5)
It is easy to read and understand (3)
It can easily determine the strengths and weaknesses (quality) of different Web Sites (3)
It considers all aspects of teaching (2)
It asks many questions that every individual might not think of. (2)
It examines the Web Sites social relevancy (2)
Looks at credibility (2)
It standardizes the evaluation method
It is detailed and concise and considers almost all aspect of designing & viewing Web Sites
It is an excellent template for Web Site creation
It provides an opportunity for critical thinking
It is clear about its objectives

12. Do you have any suggestions to improve this Web Site evaluation tool?

A page on definitions for the features teachers may not be aware of (6)
Add a section for general comments & overall quality (2)
Implement a rating system of Likert scale
Cut down the length
Emphasize content & credibility
Provide a portion for what the students think of the site
Put it on a computer

Appendix D

The Simpson Educational Web Site Evaluation Tool

The Simpson Educational Web Site Evaluation Tool©

Content Analysis

DEFINING THE WEB SITE	Strongly disagree	disagree	undecided	agree	strongly agree
The topic is clearly defined.	☐	☐	☐	☐	☐
The purpose of the web site is stated.	☐	☐	☐	☐	☐
The objectives of the web site are stated.	☐	☐	☐	☐	☐
Adequate & relevant information is providedto meet the objectives.	☐	☐	☐	☐	☐
The target audience is defined.	☐	☐	☐	☐	☐
The information is accurate.	☐	☐	☐	☐	☐
The information is current.	☐	☐	☐	☐	☐
The web site is updated frequently.	☐	☐	☐	☐	☐
Spelling, grammar, and punctuation are correct.	☐	☐	☐	☐	☐
The links are relevant and appropriate to the topic.	☐	☐	☐	☐	☐
The rationale for providing the links is explicit.	☐	☐	☐	☐	☐
The content stands alone, and is not merely a list of links to other sites.	☐	☐	☐	☐	☐

CREDIBILITY	Strongly disagree	disagree	undecided	agree	strongly agree
The identity of the author is provided.	☐	☐	☐	☐	☐
Information about the author's training and education is provided.	☐	☐	☐	☐	☐
The author's training and education are from a well-respected organization.	☐	☐	☐	☐	☐
The author provides contact information beyond an e-mail address.	☐	☐	☐	☐	☐
References and a bibliography are provided.	☐	☐	☐	☐	☐
The web site has been externally evaluated.	☐	☐	☐	☐	☐
The company producing the web site provides unbiased information.	☐	☐	☐	☐	☐
All aspects of the topic are presented.	☐	☐	☐	☐	☐

SOCIAL CONSIDERATIONS	Strongly disagree	disagree	undecided	agree	strongly agree
People of different races, ages, physical and mental abilities are included.	☐	☐	☐	☐	☐
Different groups are presented in positive and non-stereotypical ways.	☐	☐	☐	☐	☐
The web site avoids an obvious cultural bias.	☐	☐	☐	☐	☐

The social messages in the content are positive.	☐	☐	☐	☐	☐
The web site promotes inclusion and acceptance.	☐	☐	☐	☐	☐
The web site offers positive role models.	☐	☐	☐	☐	☐
The web site helps the students to develop decision-making skills.	☐	☐	☐	☐	☐

TEACHING CONSIDERATIONS	**Strongly disagree**	**disagree**	**undecided**	**agree**	**strongly agree**
The focus of the web site is educational.	☐	☐	☐	☐	☐
The web site supports the school curriculum.	☐	☐	☐	☐	☐
The reading level is appropriate for the target audience.	☐	☐	☐	☐	☐
The concepts are age-appropriate for the target audience.	☐	☐	☐	☐	☐
Definitions are provided when necessary.	☐	☐	☐	☐	☐
The content is engaging.	☐	☐	☐	☐	☐
The content is comprehensive.	☐	☐	☐	☐	☐

Style Analysis

DESIGN	**Strongly disagree**	**disagree**	**undecided**	**agree**	**strongly agree**
Graphics and animation are clear and easily interpreted.	☐	☐	☐	☐	☐
Graphics and audio are appropriate to the population.	☐	☐	☐	☐	☐
Characters are of legible size.	☐	☐	☐	☐	☐
Sound is used with purpose.	☐	☐	☐	☐	☐
The format is varied and interesting.	☐	☐	☐	☐	☐

FEATURES	**Strongly disagree**	**disagree**	**undecided**	**agree**	**strongly agree**
The web site uses interactive features.	☐	☐	☐	☐	☐
If calculations are required, they can be done on screen.	☐	☐	☐	☐	☐
Instructions are provided for all aspects of the web site.	☐	☐	☐	☐	☐
Menus are provided to make the web site user friendly.	☐	☐	☐	☐	☐

NAVIGATION	**Strongly disagree**	**disagree**	**undecided**	**agree**	**strongly agree**
Each section is labeled with a heading.	☐	☐	☐	☐	☐
Image maps are provided.	☐	☐	☐	☐	☐

TEACHING CONSIDERATIONS	Strongly disagree	disagree	undecided	agree	strongly agree
The pedagogic approach is superior using this medium.	☐	☐	☐	☐	☐
The pedagogy is innovative.	☐	☐	☐	☐	☐
Instructional strategies are appropriate.	☐	☐	☐	☐	☐
Demonstrations are clear.	☐	☐	☐	☐	☐
Minimal teacher supervision is required.	☐	☐	☐	☐	☐
There are opportunities to answer open ended questions.	☐	☐	☐	☐	☐
Students can hypothesize and experiment.	☐	☐	☐	☐	☐
The program provides a forum for student interaction.	☐	☐	☐	☐	☐
The time limits of the class are addressed by providing workable units.	☐	☐	☐	☐	☐
The material is organized and well sequenced.	☐	☐	☐	☐	☐
The web site provides criteria to assess learning.	☐	☐	☐	☐	☐
The questions in the test measure the mastery of the content.	☐	☐	☐	☐	☐
Student responses match the program objectives.	☐	☐	☐	☐	☐

Process Analysis

SYSTEM EFFICIENCY	Strongly disagree	disagree	undecided	agree	strongly agree
The homepage loads smoothly and efficiently.	☐	☐	☐	☐	☐
There is efficient transition between screens.	☐	☐	☐	☐	☐
All of the pages download.	☐	☐	☐	☐	☐
The web site run without any special software requirements.	☐	☐	☐	☐	☐
The can text stand-alone if the graphics are not loaded.	☐	☐	☐	☐	☐
Out of date links are quickly removed.	☐	☐	☐	☐	☐

TEACHING CONSIDERATIONS	Strongly disagree	disagree	undecided	agree	strongly agree
Teachers are provided with user guides for the web site.	☐	☐	☐	☐	☐

Appendix E

Describing Selected Items of the Simpson Educational Web Site Evaluation Tool

Describing Selected Items of the Simpson Educational Web site Evaluation Tool

Content Analysis

Defining the Web site

The topic is clearly defined.

This criterion asks if a topic sentence is provided. The topic sentence should provide a general idea of the content in the web site.

The purpose of the web site is stated.

This criterion examines the intent of the web site. The intent could be to educate students, to entertain students or any number of things. The purpose is one of the most important indicators for educators to decide if the web site meets their needs.

The objectives of the web site are stated.

The objectives define the different methods that the web site will use to achieve the purpose. The sum of the objectives should equal the purpose.

The target audience is clearly defined.

The target audience is the group that the web site was designed for. Educational web sites should contain a statement indicating for whom the web site was designed.

Links to external sites are relevant and appropriate to the topic.

This criterion indicates whether or not the author has selected links that complement the web site. These links may bring the user to a different level of analysis for the same topic, or may offer information on a complementary topic.

The criteria for selecting the links are made explicit.

This information will provide the user with the information required to make an informed decision as to whether or not they are interested in visiting the external web sites.

Social Considerations

Issues to these groups are in evidence.

This criterion speaks to the idea that relevant information about different groups needs to be included in the web site. The author should ensure that a diverse perspective is presented.

The Web site avoids an obvious cultural bias.

This criterion captures the idea that concepts relevant to the majority population can be stressed at the expense of the minority populations.

The Web site promotes inclusion and acceptance.
This criterion speaks to the idea that all groups are respected and treated equally within the web site.

The social messages in the content are positive.
Social messages are ideas that our society holds true.

The Web site offers positive role models.
A positive role model is someone who represents and promotes positive social ideals.

Teaching Considerations

The focus of the web site is educational.
Is the purpose of the web site to inform and educate? Are the objectives of the web site learning objectives?

The reading level is appropriate for the target audience?
This is a difficult criterion to measure since there can be different levels of literacy within the same target audience. On a subjective level the evaluator must decide if the vocabulary and the use of language is understandable by the target audience.

The content is engaging.
A user is engaged when the information is made meaningful to them. Burbles & Callister (1996) describe the difference between memorizing the information and being engaged. They indicate that "data cannot be considered information until they have been contextualized, arranged in such a way that both the significant differences and the significant relationships among them may become apparent to the intended reader" (p. 31). These relationships are not apparent if the information is not made relevant to the student.

The content is comprehensive.
The Web site must cover all aspects of a subject to be comprehensive. The purpose of the web site should indicate the level of comprehensiveness. The information should be comprehensive enough to meet the cited objectives.

Style Analysis

Navigation

Image maps are provided.
An image map is a graphic representation of the different pages of a Web site.

Teaching considerations

Pedagogy

The art or science of teaching: especially instruction in teaching methods. (Webster's)

The pedagogic approach is superior using this medium.

Web sites that take advantage of hypertext are effectively using this medium. The benefits of hypertext include the elimination of a single perspective, the ability for the user to focus their investigations on their own interests, and the ability to accommodate different personal and cultural learning styles. Hypertext allows for unstructured exploration that can support a different method of thinking in the user (Burbles & Callister, 1996). The user is not limited by the choices of the author, as they are with traditional text which offers a sequence, style and organization that is outlined by the author.

The pedagogy is innovative.

Does the Web site take advantage of the unique features of the Internet, or is it merely a textbook transposed onto a web page? These unique features include COWS, bulletin boards, e-mail, chat rooms, hypertext, Java script, and listservs.

The instructional strategies are appropriate.

Technology alone does not provide a valid learning experience. Sound teaching practices should be applied to new technology.

Process Analysis

System Efficiency

There is efficient transition between screens.

The Web site does not experience any undue delays when the user is navigating the Web site.

Appendix F

The Results of the Product-oriented Evaluation of the Web Site www.troom.com by Expert Adjudicators

The Results of the Product-oriented Evaluation of the Web site www.troom.com by Expert Adjudicators

The first column represents the number of positive responses per number of total responses. The second column represents the corresponding percentage of this relationship. The percentage in bold writing represents the number of positive responses per number of total responses for each section.

Content Analysis

78% DEFINING THE Web site

13 of 19	68%	Is the topic clearly defined?
11 of 19	59%	Is the purpose of the Web site stated?
13 of 19	68%	Are the objectives of the Web site stated?
15 of 19	79%	Is adequate and relevant information provided in order to meet objectives?
19 of 19	100%	Is the target audience defined?
18 of 19	95%	Is the information accurate?
16 of 18	89%	Is the information current?
09 of 19	47%	Was the last update recent?
19 of 19	100%	Are spelling, grammar and punctuation correct?
16 of 19	84%	Are the links relevant and appropriate to the topic?
11 of 19	59%	Are the criteria for selecting the links explicit?
18 of 19	95%	Does the content stand alone, or is it merely a list of links to other sites?

71% CREDIBILITY

16 of 19	84%	Is the identity of the author provided?
15 of 19	79%	Is the information about the author's training and education provided?
18 of 19	95%	Is this training and education from a well-respected organization?
08 of 19	42%	Does the author provide contact information beyond an e-mail address?
09 of 19	47%	Are references and a bibliography provided?
09 of 17	53%	Is the site subject to a review process?
13 of 18	72%	Does the company producing the Web site provide unbiased information?
18 of 19	95%	Are all aspects of the topic presented?
09 of 14	64%	Are there omissions due to the agenda of the group producing the Web site?

83% SOCIAL CONSIDERATIONS

15 of 19	79%	Are people of different races, ages, physical and mental abilities included?
14 of 19	74%	Are different groups presented in positive and non-stereotypical ways?
13 of 19	68%	Are issues specific to these groups in evidence?
13 of 19	68%	Does the Web site avoid an obvious cultural bias?
18 of 19	95%	Are the social messages in the content positive?
18 of 18	100%	Does the Web site promote inclusion and acceptance?

16 of 18 89% Does the Web site offer positive role models?
17 of 18 94% Does the Web site help the students to develop decision-making skills?

91% TEACHING CONSIDERATIONS

19 of 19 100% Is the focus of the Web site educational?
08 of 11 73% Does the Web site support the school curriculum?
11 of 12 92% Is the reading level appropriate for the target audience?
12 of 12 100% Are the concepts age-appropriate for the target audience?
17 of 18 94% Are definitions provided when necessary?
06 of 08 75% Is the content engaging?
17 of 18 94% Is the content comprehensive?

Style Analysis

71% DESIGN

19 of 19 100% Are graphics and animation clear and easily interpreted?
14 of 19 74% Are graphics and audio appropriate to the population?
09 of 19 47% Does the visual display occupy most of the screen?
18 of 19 95% Are the characters of legible size?
03 of 17 18% Is the sound used with purpose?
17 of 19 90% Is the format varied and interesting?

58% FEATURES

11 of 18 61% Does the Web site use interactive features such as CGI scripts & conferencing?
07 of 17 41% If calculations are required, can they be done on screen?
03 of 16 19% Are bookmarks and notepads provided for the students?
01 of 06 17% Is clipping information and automatic citations possible?
02 of 09 22% Are students denied access to confidential or inappropriate information?
13 of 19 68% Are instructions provided for all aspects of the Web site?
16 of 19 84% Does the student control the rate of everything presented?
18 of 19 95% Are menus provided to make the Web site user friendly?

78% NAVIGATION

19 of 19 100%Is each section labeled with a heading?
10 of 13 77% Are image maps provided?
04 of 07 57% Is a search engine available to the user?
02 of 06 33% Does the search engine index the entire Web site?

66% TEACHING CONSIDERATIONS

02 of 05 40% Is the pedagogic approach superior using this medium?
04 of 05 80% Is the pedagogy innovative?
08 of 08 100% Are instructional strategies appropriate?
18 of 18 100% Are demonstrations clear?
18 of 19 95% Is minimal teacher supervision required?
10 of 18 56% Is there opportunity to answer open ended questions?
08 of 19 42% Can students hypothesize and experiment?

08 of 09 89% Does the program provide a forum for student interaction?
10 of 18 56% Are the time limits of the class addressed by providing workable units?
18 of 19 95% Is the material organized and well sequenced?
10 of 18 56% Does the Web site provide criteria to assess learning?
09 of 18 50% Do the questions in the test measure the mastery of the content?
03 of 16 19% Do student responses match the program objectives?
06 of 09 67% Is feedback provided that is positive, informative and timely?

Process Analysis

42% INPUT CONSIDERATIONS
02 of 09 22% Does the Web site allow users to correct input errors?
04 of 09 44% Does the Web site accept abbreviations?
03 of 07 43% Are minor variations in spelling and grammar allowed?
02 of 07 29% Are control keys used consistently?
06 of 09 67% Are both mouse and keyboard control allowed?

80% SYSTEM EFFICIENCY
16 of 16 100% Does the homepage load contiguously?
06 of 19 32% Can the startup screen be bypassed?
18 of 19 95% Is there efficient transition between screens?
17 of 19 90% Can the Web site run without any special software requirements?
18 of 19 95% Can the text stand-alone if the graphics are not loaded?
09 of 17 53% Are out of date links quickly removed?
18 of 19 95% Do all of the pages download?

40% TEACHING CONSIDERATIONS
06 of 19 32% Can the teacher change or add content?
01 of 08 13% Can the Web site store the test scores of a number of students?
04 of 08 50% Are the results easily accessible?
06 of 08 75% Are teachers provided with user guides for the Web site?

27% SUPPORT
03 of 19 16% Is a help function available to users at all times?
08 of 19 42% Are the materials required to run the site defined?
04 of 18 22% Are any support materials required, and if so, are they provided?

Appendix G

Protocol for the Completion of the Simpson Educational Web Site Evaluation Tool

Protocol for the Completion of the Simpson Educational Web site Evaluation Tool

1. The Simpson Educational Web site Evaluation Tool will be introduced.

2. The format of the evaluation tool is a Likert scale. The evaluator is to indicate the degree to which they agree with the criterion statement by placing a check mark in the appropriate box.

3. If the evaluator does not understand the question please place a question mark in the undecided box.

4. If the evaluator does not feel they have competency in any of the criterion statements, please do not check any of the boxes.

5. Please place an asterisk beside the criteria that you feel you have expertise in.

6. Please fill in the general comment section with an overall impression of the Web site, as well as any information that you feel the checklist did not capture.

Appendix H

Form Completed by End-users in Pilot Study 1

Form Completed by End-users in Pilot Study 1

The _____ Web site Evaluation Tool

Authors _____

Please answer the following questions:
Name 10 things you like about the Web site
Name 5 things you do not like about the Web site
List what you would do to improve the Web site

Things to consider when making your Web site Evaluation Tool

Think about the Web site in three different parts. The content, the style and the process.

Content: ***The information located in the Web site.***
Style: ***The way that the information is presented.***
Process: ***The way that the Web site works.***

What kinds of questions do you need to ask about the content to decide if the content is good or not? ***Hints: What are you graded on when you hand things in to your teacher? What makes a book really good? Why do you have textbooks?***

What kinds of questions do you need to ask about the style? ***Hints: What are the different ways that information is presented? What makes a Web site really cool? What makes a Web site boring? What is the difference?***

What kinds of questions do you need to ask about the process? What makes a Web site good? ***Hints: Do you have to wait a long time for the web page to load? What do you use to move around on the Web site? What happens when you aren't able to move around on the Web site?***

Look at the Web site www.bodymatters.com. Use the criteria that you came up with to evaluate this Web site. Having looked at this Web site, what are some of the criteria that you think you missed? What else do you think is an important part of an educational Web site?

Appendix I

The Results of the Product-oriented Evaluation of the Web Site www.troom.com by End-user Adjudicators

The Results of the Product-oriented Evaluation of the Web site www.troom.com by End-user Adjudicators

The following two lists were generated from the questions: i) Name 10 things you like about the Web site ii) Name 5 things you do not like about the Web site. The student's responses have been grouped into the categories of content, process and style.

Things the end-users liked:

Content

Good source of information (10)*
helps people with questions they don't want to ask(2)
Interesting (2)
Educational (2)
Liked the activities (2)
Updated every month (2)
You can get good advice
Has tips on personal care
Helps you to not be embarrassed

specific activities judged good

Games (11)
Funny fill-ins (8)
Diaries (6)
Music reviews (6)
Sports & fitness(4)
The calendar (4)
The phone (2)
Penpals (4)
Laugh out loud
What's new
Stories
Lookin' good
Tina net
The travel room
Quizzes
Information about stars
Comics
My room

Things the end-users disliked

Content

Too much girl stuff (10)
Should have more guy stuff (3)
Too personal (3)
Too commercial, too many sponsors (2)
Don't like the subject matter (2)
Some of the optional sites are redundant
Don't like it when you show the pads
There were words we don't understand
It's embarrassing
There are not enough games
This is not a good site for kids

Specific activities judged bad

The Questions
Reflections
Diaries
The calendar
Top drawer

* The number in brackets indicates the number of students with the same response

Process

It is easy to get around (5)
Doesn't take too long to load because of too many graphics

Process

It is difficult to get around (2)
It's too complicated

Style

Good graphics (15)
Colorful (11)
Funny (8)
Lots of choice (6)
Good Variety (5)
Easy to use (3)
The Web site is realistic (2)
Nice backgrounds (2)
Good layout (2)
Liked the index (2)
Fun, creative and eye catching
You can play on it
Important points are underlined
Titles are in bold letters
Nice font
Easy to read

Style

Pictures are too graphic (4)
The room is a mess (2)
It is boring at times (2)
It is too colorful
The teddy bears
There are too many choices

The End-users suggestions for improving the Web site

Try to make it for both genders, not just for girls. (3)
Needs better music (2)
Needs a warning screen to let boys know they are going to have to look at tampons
Cut back on the graphic pictures
Add more games
Use frames
Provide a picture of Tina so people know who's advice they are asking for
Have a chatroom
Have a comment page
Make it easier to get from one program to another
Needs more Canadian content

Appendix J

Pretest and Posttests for the End-user Performance Appraisal

Guide Book
Puberty and Menstrual Health

Activity Sheet 1

Facts and Fables

We undergo many changes in our lives -- some planned for, others unexpected. One time of change for everyone comes in the early teen years, when our bodies change and our goals, expectations and views of life take new shape. What do you feel and know about the physical changes that occur at puberty? Below are 25 statements about puberty, menstruation and menstrual protection. Find out how you feel about these topics by putting a check in the appropriate box.

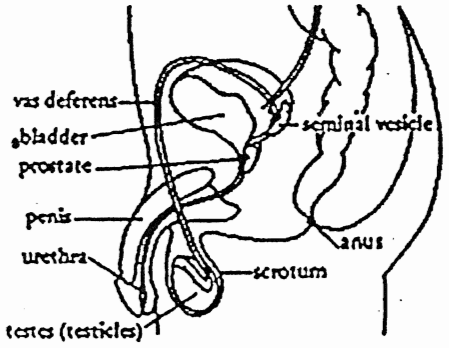
		True	False	Don't Know
1.	I already know all I need to about puberty and menstruation.			
2.	All females have menstrual cramps during their period.			
3.	Boys don't undergo changes at puberty.			
4.	Menstruation should have no effect on a female's ability to work.			
5.	I don't like the changes that are happening to my body right now.			
6.	Males should know about menstruation.			

7.	You can't go swimming when you have your period.			
8.	Eating a nutritious, well-balanced diet can help prevent some menstrual discomforts.			
9.	Young girls can use menstrual tampons.			
10.	Females shouldn't exercise during their menstrual period.			
11.	I'm embarrassed to talk with my parents about the changes I am experiencing.			
12.	Females are more emotional during their period.			
13.	I have already learned about all the changes that take place during puberty.			
14.	During puberty, boys and girls can become parents.			
15.	Boys undergo puberty at a different age from girls.			
16.	There is nothing you can do for menstrual discomfort.			
17.	Having your period keeps you from doing the things you like to do.			
18.	Boys can tell when a girl has her period.			
19.	Careful, thorough cleansing of your face can help prevent skin blemishes.			
20.	All boys have nocturnal emissions.			
21.	I have talked with my friends about the changes in our bodies.			
22.	Tampons are comfortable to use.			
23.	People shouldn't talk about menstruation.			
24.	During puberty boys and girls sweat more.			
25.	After girls begin to menstruate, it is unusual to miss a period.			

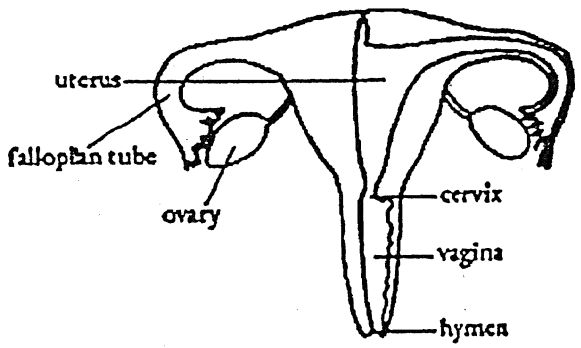
From Boy to Man

		True	False	Don't Know
1.	Sperm and semen are the same thing.			
2.	The hormone testosterone causes a boy's voice to deepen.			
3.	The hormone testosterone causes hair to grow on the body, face and pubic area.			
4.	The testicles are very delicate and should be supported during sports activities.			
5.	It is possible for semen and urine to mix together and pass from the body at the same time.			
6.	The "shrinking" of testicles in cold water is not a cause for concern.			
7.	An erection occurs when blood vessels in the penis are filled with an extra supply of blood.			
8.	Ejaculation can occur when there is a build-up of excess seminal fluid.			
9.	A boy's breasts can get bigger and become sore during puberty.			
10.	Men usually stop producing semen when they are around 60 years old.			
11.	Boys who have pimples or acne have more male hormones than other boys.			
12.	During puberty a boy's shoulders broaden and he begins to grow taller.			
13.	Circumcision is an operation that removes the foreskin of the penis.			
14.	Nocturnal emissions are caused by eating too much fried food.			
15.	New thoughts and feelings accompany the physical changes of puberty.			
16.	The size of a boy's penis determines how masculine he will be.			
17.	Sexual excitement can cause body odor.			
18.	The voice deepens when the larynx (voice box) gets larger and the vocal cords get longer.			
19.	Boys continue to grow until they are 20 years old.			
20.	At puberty, a boy is capable of becoming a father.			

_____ Pituitary Gland	1. Hormone-producing gland located at the base of the brain.
_____ Hypothalamus	2. External sac holding testes.
_____ Scrotum	3. Sperm-producing glands.
_____ Prostate	4. Narrow tubes that carry sperm from testes.
_____ Testes(testicles)	5. A fluid producing gland.
_____ Urethra	6. Narrow tube through which urine or semen pass through the penis.
_____ Penis	7. Organ through which semen or urine leaves the body.
_____ Vas deferens	8. A thick fluid containing sperm.
_____ Semen	9. Area of the brain that controls the pituitary gland.
_____ Seminal Vesicles	10. Two sacs that produce a thick fluid that carries the sperm.



_____ Endometrium	1. Hormone-producing gland located at the base of the brain.
_____ Uterus	2. Egg-producing gland.
_____ Ovary	3. Tube through which egg or ovum travels to the uterus.
_____ Vagina	4. Organ within which a baby can develop.
_____ Menstruation	5. The lining of the uterus.
_____ Fallopian tube	6. Passageway through which menstrual flow leaves the body.
_____ Pituitary gland	7. Periodic shedding of the lining of the uterus.
_____ Hymen	8. Flexible fold of tissue at entrance of the vagina.
_____ Hypothalamus	9. Lower portion of the uterus which protrudes into the vagina.
_____ Cervix	10. Controls the pituitary gland.



F = Fact M = Myth

Even today, some myths linger on. See if you can spot the myths in the following statements. Please: circle the correct letter

Fact or Myth?

F M 1. Once a girl has had her first period, she can become pregnant.

F M 2. Girls are often sick during their period.

F M 3. Females are born with unripened eggs in their ovaries.

F M 4. Girls should avoid sports when they're menstruating.

F M 5. If a woman misses her period, she is pregnant.

F M 6. A girl cannot get pregnant if she has sex only once in awhile.

F M 7. A girl can lose her virginity by using a tampon.

F M 8. It's not safe to wash your hair or take a bath during your period.

A. Menopause	____1. Nausea, cramps or dizziness around the time of menstruation
B. Premenstrual syndrome	____ 2. Lack of menstrual flow
C. Dysfunctional uterine bleeding	____ 3. Feelings of depression, irritability, headaches, bloating, temporary weight gain, sore breasts, joint pain and general tiredness
D. Hysterectomy	____ 4. Excessive bleeding betewen or during periods
E. Symptoms of Toxic Shock Syndrome	____ 5. A yellowish or colored discharge between periods that causes itching, odor or a burning sensation
F. Amenorrhea	____ 6. A sudden high fever, vomiting, diarrhea, muscle aches, a rash that looks like a sunburn, dizziness, fainting or near fainting when standing
G. Abnormal vaginal discharge	____ 7. Surgical removal of the uterus
H. Dysmenorrhea	____ 8. The permanent ending of the menstrual cycle
I. Osteoporosis	____ 9. The thinning and deterioration of bone due to age and loss of estrogen

Appendix K

The End-user Educational Web Site Evaluation Tool

The End-User Educational Web Site Evaluation Tool

Content Analysis

	Yes	undecided	No
Is the web site a good source of information?	☐	☐	☐
Has the web site taught you things you did not already know?	☐	☐	☐
Does the web site have words you do not understand?	☐	☐	☐
Does the web site addresses topics you do not understand?	☐	☐	☐
Is this a good web site for people your age?	☐	☐	☐
Are people from different cultures represented in the web site?	☐	☐	☐
Do you have questions or concerns that are not answered by the web site?	☐	☐	☐

Style Analysis

	Yes	undecided	No
Is the web site interesting?	☐	☐	☐
Are the graphics clear?	☐	☐	☐
Are the graphics suitable for people your age?	☐	☐	☐
Is the web site well laid out?	☐	☐	☐
Does the web site offer choice and variety?	☐	☐	☐
Are the activities enjoyable?	☐	☐	☐
Does the web site have enough interactive features?	☐	☐	☐
Is the web site boring?	☐	☐	☐
Are instructions provided when you need them?	☐	☐	☐

Process Analysis

	Yes	undecided	No
Is it easy to move around the web site?	☐	☐	☐
Do the pages of the web site load quickly?	☐	☐	☐
Do all of the pages download?	☐	☐	☐
Did you get lost in the web site?	☐	☐	☐

Appendix L

Verbal Protocol Applied to the End-users

Verbal Protocol Applied to the End-users During Evaluation of the Web site www.bodymatters.com and the Completion of a Learning Appraisal

Location: Brock University, Education Building

Pre-test (9:35-9:50)

When I say to begin please complete the activity sheets located on the desk in front of you. Please complete the activity sheets without help from the Web site or your fellow classmates.

You will find some of the questions do not have a correct or incorrect answer. Your opinions will not be marked, so please feel free to answer these questions as you see fit.

If you have any questions please feel free to ask us. Please keep in mind that we cannot give you the answer to a question.

You will have 15 minutes to complete the task. When you have finished completing the activity sheets please put your hand up so that we can collect them. Do you have any questions? Please begin.

Intervention (9:50-10:35)

Now that you have completed the activity sheets you may turn on the monitor in front of you. The Web site www.bodymatters.com has been downloaded for you.

Please take the time to examine the Web site carefully. You will have 45 minutes to read the information on the Web site.

Remember that this is a test of the Web site and not of your ability. I want to make sure that any improvements that you make are because you have found the answers in the Web site. Please avoid the Answers portion of the Web site. I will not be able to include the results of any student who looks at the Answer page in my experiment. Do you have any questions?

Post-test

Please complete the activity sheets in front of you. Do you have any questions?

Break (10:45-10:55)

End-user Evaluation

Now that you have had the opportunity to use the Web site as it was designed to be used, I would like to ask your opinion of it. The paper in front of you is called the End-user Evaluation Tool. The End-user is you....the population that the Web site was designed for. Let me walk you through the first question to demonstrate how the evaluation tool is to be completed. "Is the Web site a good source of information?" If you think it is, please check yes, if you think it is not please check no. If you cannot decide between yes and no, please check the undecided box. The second sheet is marked "comments". On this sheet please feel free to write down any comments you have about the Web site. You will have 20 minutes.

Appendix M

Expert Adjudicator Information Form

Expert Adjudicator Information Form

Please indicate the following where relevant to your status as an "expert" in puberty education, pedagogy, or Internet.

Formal Education:

_____

_____

_____

_____

Informal Education:

_____

_____

_____

_____

Employment Experience:

_____

_____

_____

_____

_____

Would you consider yourself to be at the beginner, intermediate, or advanced level of experience for navigating web sites? _____

Please indicate the length of time spent examining the Web site and completing the Simpson Educational Web Site Evaluation Tool.

_____